



HORIZON 2020
THEME SC5-2017



European Climate Prediction system
(Grant Agreement 776613)

European Climate Prediction system (EUCP)

Deliverable D2.4

Identification of optimal subsets of future climate realisations and provision of event set of present and future extreme events

Deliverable Title	<i>Identification of optimal subsets of future climate realisations and provision of event set of present and future extreme events</i>	
Brief Description	<i>Overview of methods to determine optimal subsets of CMIP5/6 or downscaled model subsets and methods to create event sets of present and future extreme events.</i>	
WP number	2	
Lead Beneficiary	<i>Karin van der Wiel, KNMI</i>	
Contributors	<i>Karin van der Wiel, Erik van Meijgaard, KNMI Marianna Benassi, Silvio Gualdi, Giusy Fedele, CMCC Anna Merrifield, Lukas Brunner, ETHZ Carol McSweeney, Tamzin Palmer, MetOffice Renate Wilcke, SMHI</i>	
Creation Date	<i>17/02/2021</i>	
Version Number	<i>2.0</i>	
Version Date	<i>22/11/2021</i>	
Deliverable Due Date	<i>30/11/2021</i>	
Actual Delivery Date	<i>30/11/2021</i>	
Nature of the Deliverable	R	<i>R – Report</i>
		<i>P - Prototype</i>
		<i>D - Demonstrator</i>
		<i>O - Other</i>
Dissemination Level/ Audience		<i>PU - Public</i>
		<i>PP - Restricted to other programme participants, including the Commission services</i>
		<i>RE - Restricted to a group specified by the consortium, including the Commission services</i>
	X	<i>CO - Confidential, only for members of the consortium, including the Commission services</i>

Version	Date	Modified by	Comments
1.0	11 Oct 2021	Deliverable contributors	For EUCP internal review
2.0	22 Nov 2021	Deliverable contributors	WP version submitted to EUCP board

Table of contents

1. Executive Summary	4
2. Project Objectives	5
3. Detailed Report	5
3.1. Introduction	5
3.2. Science part 1 - Pruning the CMIP5/6 Family Tree using ClimWIP (ETHZ)	6
3.3. Science part 2 - CMIP6 sub-selection based on an assessment of key performance criteria (MO)	12
3.4. Science part 3 - Towards subselection based on circulation changes: changes in Jet Latitude and Strength and the Summer NAO (MO)	18
3.5. Science part 4 - Selecting regional simulations tailored to user's needs and setting small (CPRCMs) and large (GCMs) ensembles into uncertainty context (SMHI)	24
3.6. Science part 5 - Quantifying the minimum number of ensemble members (CMCC)	31
3.7. Science part 6 - Pseudo-global warming experiments to create event sets of future weather (KNMI)	35
3.8. Science part 7 - Selecting weather events from initial condition large ensembles (KNMI)	39
3.9. List of related EUCP publications	44
3.10. Discussion	44
4. Lessons Learnt and links built	45
5. References	46
A1 - Appendix to Science part 1	51
A2 - Appendix to Science part 2	63
A3 - Appendix to Science part 4	67
A4 - Appendix to Science part 6	69
A5 - Appendix to Science part 7	94

1. Executive summary

This deliverable reports on EUCP work in WP2 T2.3 regarding methods of sub-selection in climate science. Data reduction without losing information, the ‘distillation’ of data in IPCC terms, is highly relevant for example for downscaling purposes or impact modelling applications that cannot take the full CMIP6 ensemble as input. In such cases an informed choice of which models, ensemble members or periods to use will help to better represent the spread of driving models in experiments, thereby making results and outcomes more usable and useful.

Furthermore this deliverable reports on methods for the development of event sets of future weather. Information on future weather conditions, in addition to widely available information on future mean climatic conditions, is relevant as it more directly relates to (societal) vulnerability and impacts, and thus can more directly inform users or policy makers.

The report describes the development of several distillation methods, through seven highlights named ‘science parts’. These showcase practical examples of different selection methods and methods for creating future weather event sets, including examples of their results. Overview of the science parts:

1. The ClimWIP tool weighs or selects CMIP5/6 climate model runs based on independence and performance. Open-access software is available to allow a user to find a representative set of high performing, independent models for a task.
2. A qualitative sub-selection method, ranking models from ‘implausible’ to ‘satisfactory’ based on regional model performance in key climate processes. Users can take this information on board when making their model choices.
3. Selecting members based on large scale characteristics. An exploration of the spread in changes explained by large scale characteristics using the Summer NAO and the position and strength of the Jet Stream.
4. Model selection method using singular vector decomposition and hierarchical clustering with the aim to create a smaller model subset that retains the variability of user relevant climate information.
5. A method to investigate internal variability and how many ensemble members are required to adequately sample the complete distribution of variability. This is useful for both end users (e.g. to define input for a tailored impact study) and climate scientists (e.g. in experimental design).
6. A regional climate modelling method based on pseudo-global warming experiments to create event sets of future weather similar to observed weather. An example for a set of future droughts like the 2018 drought is provided.
7. A sub-selection method based on initial-condition large ensemble data to create event sets of future weather similar to observed weather. An example for future droughts like the 2018 drought is provided.

A short wrap-up is provided at the end, in which we note similarities between the methods and studies presented. Furthermore we provide some advice for future users of selection methods and for scientists building on the work of EUCP.

2. Project objectives

These deliverables have contributed to the following EUCP objectives (Description of Action, Section 1.1):

No.	Objective	Yes	No
1	Develop an ensembles climate prediction system based on high-resolution climate models for the European region for the near-term (~1-40 years)	Yes	
2	Use the climate prediction system to produce consistent, authoritative and actionable climate information	Yes	
3	Demonstrate the value of this climate prediction system through high impact extreme weather events in the near past and near future	Yes	
4	Develop, and publish, methodologies, good practice and guidance for producing and using EUCP's authoritative climate predictions for 1-40 year timescales	Yes	

3. Detailed report

3.1 Introduction

This deliverable outlines the work of EUCP WP2 T2.3 regarding methods of sub-selection in climate science. Such (sub-)selection can refer to the selection of specific climate models from a multi-model ensemble of climate models (e.g. CMIP6, EURO CORDEX), the selection of specific climate model runs from a single model initial-condition large ensemble (Deser et al. 2020), or specific climatic events from a longer climate model run. Many reasons for performing such selections exist, in this section we introduce some of these reasons.

Climate change impacts societies and natural systems around the world, though its impacts and effects vary from region to region and from sector to sector. Actionable climate information therefore depends on the specific region considered, but also on the precise user that is considered or the question that is asked. Chapter 10 of the recently published IPCC report (IPCC 2021) describes the process of constructing 'user-oriented climate information'. Information from observational data, output from climate models, the scientific literature and climate experts, must be combined and 'distilled', to add value and increase context-relevance. The methods for this 'distilling' of data, data reduction without losing information, are topics of ongoing research. Their importance is well-known though, IPCC notes "*Data, either from observations or models, is in general not inherently information, but may contain relevant information if interpreted appropriately (Hewitson et al., 2017).*"

In this report we discuss work on several sub-selection/distillation methods, sections 3.2 to 3.8 are highlights of the scientific progress made within this task. They show practical examples of different selection methods and (an example of) their results. For example, whilst the CMIP5 and CMIP6 model

ensembles include a wealth of climate data, model dependencies, model biases, or simply the large number of models make sub-selection necessary. Informed sub-selection, taking into account quality, (in)dependence, and the range of uncertainty, is much preferable over random sub-selection. A biased sub-selection of CMIP5 GCMs to drive EURO-CORDEX RCMs might for example be the cause of the RCM ensemble missing the upper part of the climate change uncertainty range (Boé et al. 2020). The ClimWIP tool (section 3.2) or the qualitative performance assessment (section 3.3) were developed with such applications in mind, and provide guidelines for optimized sub-selection based on user requirements.

Section 3.4 investigates the possibility of data distillation through developing understanding of key circulation characteristics, how they change in response to climate change and how they influence regional climate. This can potentially lead to the identification of a number of axes along which uncertainty maximizes and from which model sub-selection can be done. The resulting subset of models can then be used for storyline development of future change (Zappa and Shepherd, 2017; Shepherd et al. 2018). Section 3.5 reflects on downscaling experiments, and puts the outcomes of high-resolution downscaling ensembles into the context of the GCM/RCM ensembles that they are based on. This also show Section 3.6 develops methods to investigate internal variability and how many ensemble members are required to adequately sample the complete distribution of variability.

Finally, sections 3.7 and 3.8 describe two independent methods for the construction of future extreme weather events. Tangible examples of future weather, in the cases discussed here linked to recent observed weather events, can be very useful for informing adaptation policy or stress tests. In section 3.7 a set of Pseudo Global Warming regional model experiments is performed to assess the development of a drought like 2018 in a warmer climate. This approach is complemented in section 3.8 by a sub-selection approach using initial-condition large ensemble data. The methods developed can be applied towards the EUCP aim of providing event sets of future weather.

In summary, this deliverable aims to describe and evaluate different methods of (climate data) sub-selection, with the aim to contribute to more actionable user-oriented climate change information. Some of the highlighted studies have been published in the peer-reviewed scientific literature (listed in Section 3.9), others are undergoing peer review at time of writing or are still being developed further before publication. We have aimed to keep the main report concise, by writing concise summaries of the scientific results. Where appropriate appendices including more detailed methodological descriptions or the full results are provided.

3.2 Science part 1 - Pruning the CMIP5/6 Family Tree using ClimWIP (ETHZ)

MOTIVATION -- There is a growing consensus among climate scientists that our research must reach a broad user community; stakeholders, policymakers, and interdisciplinary researchers seek accessible, actionable climate information based on the most up-to-date, rigorously vetted climate model output available. Delivering this information presents a challenge. Which models realistically simulate a region, season, or process of interest? Are newer model versions demonstrably superior to their predecessors? If so, is it necessary, for example, to update the global models that drive regional models each time new simulations are published? While diagnostic studies may document systematic

physical climate biases or other model quality issues, many are built to advise but not actively guide users in the model selection process.

The EUCP project is uniquely positioned to actively guide users to the best European climate information available. Model selection logically follows from the performance and independence weighting strategies developed in WP2 (Brunner, et al. 2020a) to provide constrained estimates of mid- and end-of-century regional European temperature and precipitation change. In this deliverable, we illustrate how one WP2 weighting strategy, Climate model Weighting by Independence and Performance or ClimWIP (Brunner et al. 2019, Brunner et al. 2020b), can be utilized to select simulations from distinct climate model families within the CMIP5/6 archives that project realistic mid-century Central European summer climate change. The ClimWIP sub-selection protocol lessens the burden of updating models generation-to-generation by allowing users to generate representative model subsets from multi-generational archives with minimal computational burden.

APPROACH -- To select high performing, reasonably independent climate models best suited for applications requiring realistic estimates of mid-century Central European summer surface air temperature (SAT) and precipitation (PR) change, we first compute aggregated distance from observation, D_i , for each member of a combined CMIP5/6 archive using ClimWIP. The CMIP5/6 archive consists of historical + RCP8.5 projections for CMIP5 and historical + ssp585 projections for CMIP6. The two future scenarios are comparable though not identical (Tebaldi et al., 2020); for discussion of rationale and lists of all members used, please refer to Appendix A1.2. ClimWIP is the performance and independence weighting scheme first applied in Knutti et al. (2017), which has brought together model evaluation and notions about commonalities in model development to provide justifiable representations of uncertainty in multi-model ensemble projections of future climate (Sanderson et al. 2015, 2017). It has since been integrated into the Earth System Model Evaluation Tool (ESMValTool; Righi et al. 2020) for general open use as part of the EUCP effort towards methodological open access ([Climate model Weighting by Independence and Performance \(ClimWIP\) — ESMValTool 2.3.0 documentation](#)).

Aggregated distance from observation, D_i , is computed as the sum of area-weighted root-mean-square-errors (RMSE) between CMIP5/6 members and observations for the following “performance predictors” :

- Annual average European SAT Climatology; 1950-1969
- Annual average European SAT Climatology; 1995-2014
- JJA average Central European Station PR Climatology; 1995-2014
- JJA average Central European SREX (CEU) SWCRE Climatology; 2001-2018
- Annual average North Atlantic SST Climatology; 1995-2014
- Annual average Southern Hemisphere midlatitude SWCRE Climatology; 2001-2018

SWCRE is the difference between all and clear sky downwelling shortwave radiation at the surface (shortwave cloud radiative effect) and SST is sea surface temperature. All fields are remapped onto a $2.5^\circ \times 2.5^\circ$ latitude–longitude grid (Brunner et al. 2020b) prior to the computation of RMSE. RMSEs for each predictor are then normalized by the predictor mean before being added together. For more information about predictor choices, the observational datasets used, and other methodological details, please refer to Appendix A1.3 and 1.4.

Upon computing D_i , we commence with the first selection round: to select the highest performing (most suitable for the task of simulating Central European summer climate change) member from each “model ensemble” within the CMIP5/6 archive. We define “model ensemble” as an initial condition and/or perturbed physics ensemble where members share the same model name. From each model ensemble, the member with the lowest value of D_i (closest to observed) is selected to become a branch on the CMIP5/6 family tree shown in Fig. 1. It is notable that the highest performing member of each model ensemble (as selected by ClimWIP) is often not the first member. ClimWIP offers an alternative to the approach of simply selecting “member 1” when multiple members of a model exist (Karlsson & Svensson, 2013; Liu et al., 2012; Sillmann et al., 2013); an alternative that ensures all models are represented in the best possible light.

To formulate the CMIP5/6 family tree, a hierarchical clustering approach (Bar-Joseph et al., 2001; Müllner, 2011) is taken using ClimWIP’s inter-model distance metric, S_{ij} , which is used to determine to what degree a model is a dependent “close relative” of another model. S_{ij} is also a cosine latitude weighted RMSE-based metric, computed between all model pairs, based on predictors found to identify initial condition ensemble members as members of the same model and to assign degrees of dependence to closely related models (Merrifield et al. 2020). Models are grouped using average-linkage criteria; distances to existing clusters are calculated using cluster average S_{ij} . In the resulting family tree dendrogram, models with many close relatives (more branches left to right) are grouped at the top of the page and models with fewer similar counterparts attach to the tree nearer to the bottom of the page (Fig. 1). For more information about construction of the family tree, please refer to Appendix A1.5.

Because the D_i of each model is known, we are able to “prune” the lower performing branch from each stem to reveal subsets of independent CMIP5/6 members. Users can refer to the tree to select the highest performer from a particular model family, for example, HadGEM3-GC31-MM from the family of models, listed in Fig. 1 in red, with UK Met Office Unified Model and Hadley Centre Global Environment Model atmospheres (Andrews et al., 2019). It is of note that CMIP6 models tend to, but don’t always, perform better than their CMIP5 predecessors. CMIP5’s GFDL-CM3 is the second highest performer in the set, behind only CMIP6’s AWI-CM-1-1-MR. The ten highest performers (listed in the inset of Fig. 1) tend to come from the more highly branched upper portion of the tree, indicating that they are close relatives of one another. For use cases such as regional modelling, a range of high performing independent models is often more valuable than a set of high performing very similar models. Because of this, we also recommend a set of independent high performers, listed in the inset of Fig. 1, that we select through the minimization of a performance/independence cost function.

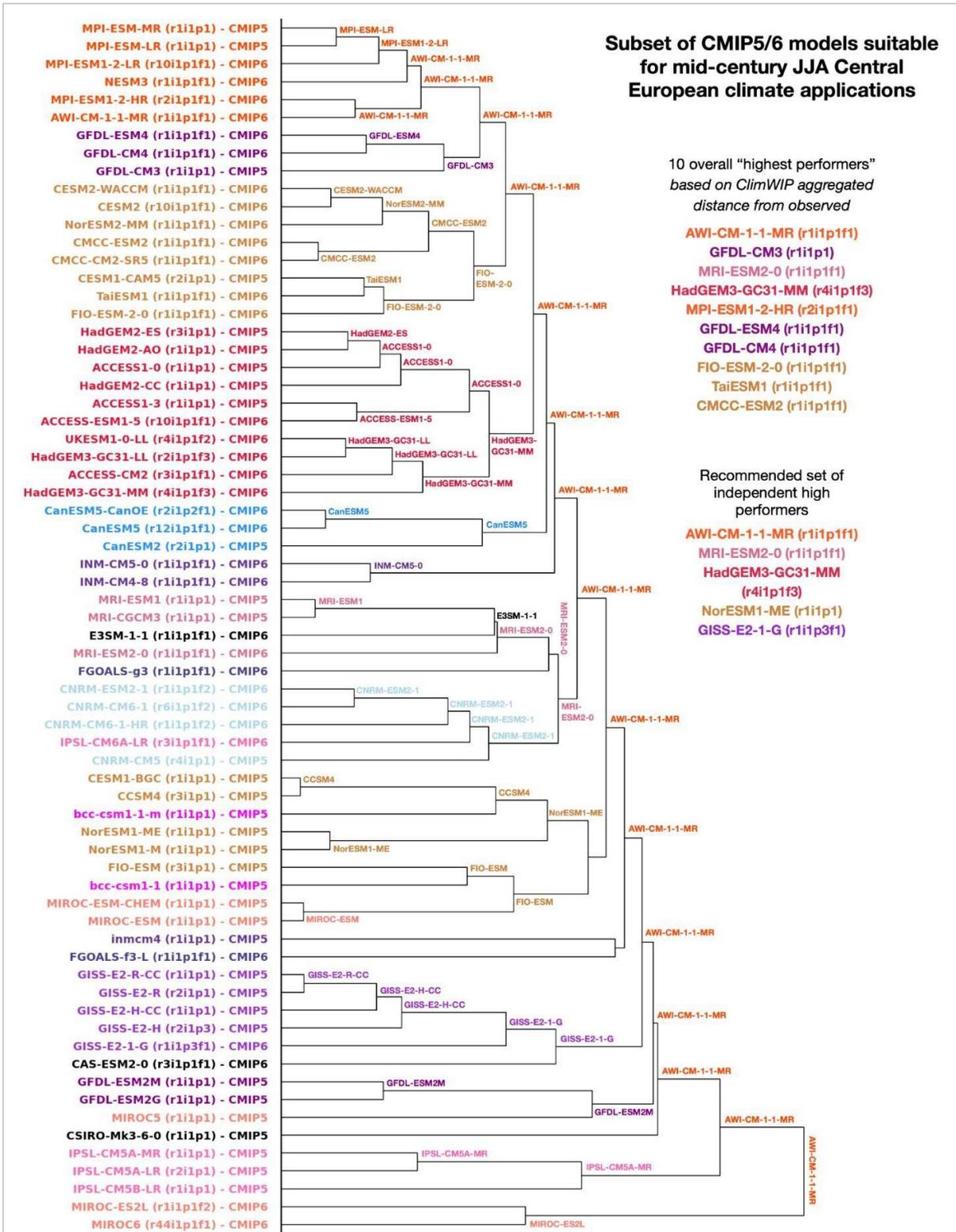


Figure 1. CMIP5/6 family tree created using *ClimWIP* independence predictors (following Merrifield et al. (2020); annual average global land temperature climatology and annual average Northern Hemisphere sea level pressure climatology;1950-2014). Each CMIP5/6 model is initially represented by its highest performer. Following each branch to the right, the higher performer of the pair continues while the lower performer is “pruned”. The 10 highest performers for mid-century JJA Central European climate applications and a recommended set of 5 independent high performers are listed in the inset.

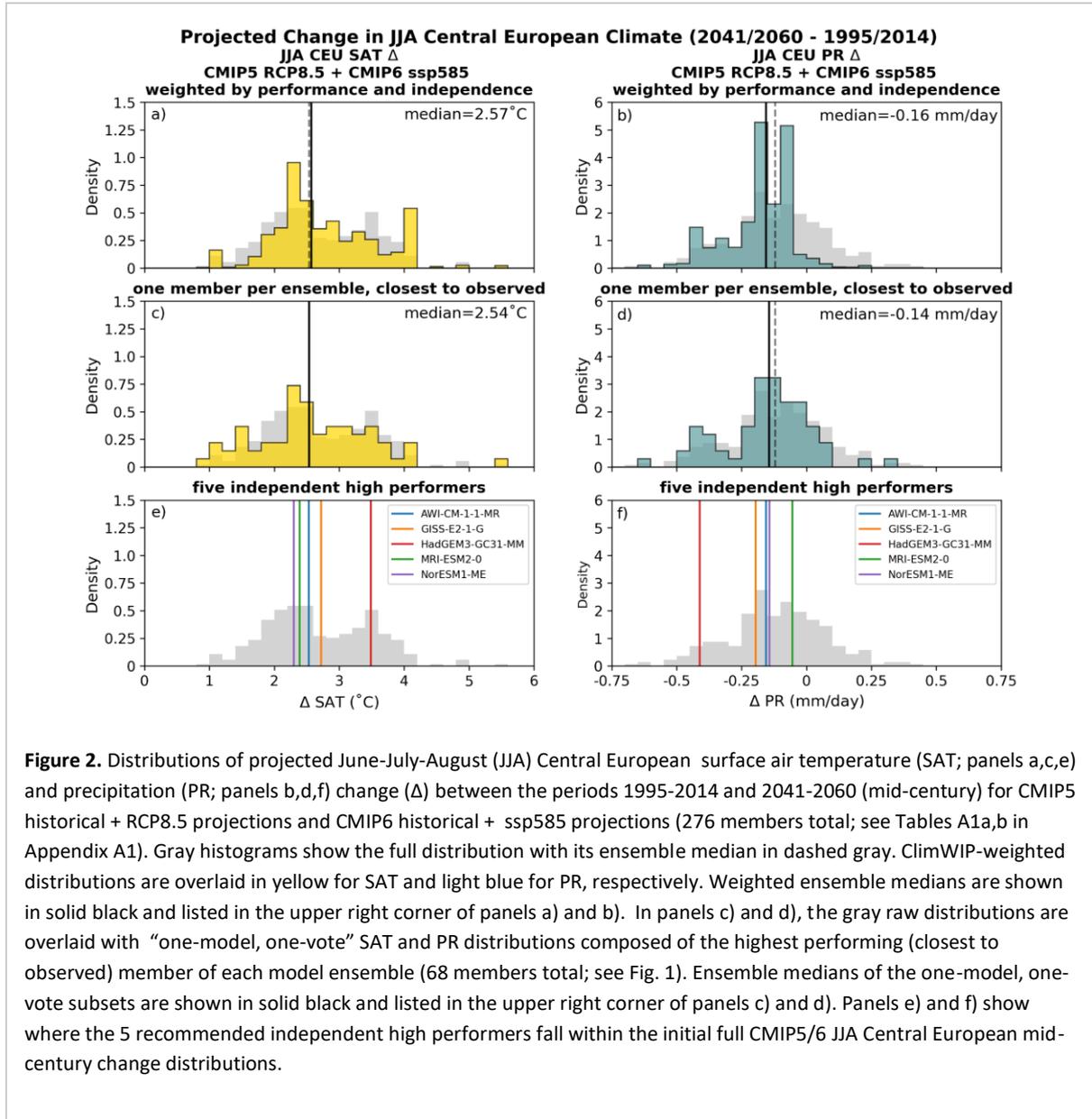


Figure 2. Distributions of projected June-July-August (JJA) Central European surface air temperature (SAT; panels a,c,e) and precipitation (PR; panels b,d,f) change (Δ) between the periods 1995-2014 and 2041-2060 (mid-century) for CMIP5 historical + RCP8.5 projections and CMIP6 historical + ssp585 projections (276 members total; see Tables A1a,b in Appendix A1). Gray histograms show the full distribution with its ensemble median in dashed gray. ClimWIP-weighted distributions are overlaid in yellow for SAT and light blue for PR, respectively. Weighted ensemble medians are shown in solid black and listed in the upper right corner of panels a) and b). In panels c) and d), the gray raw distributions are overlaid with “one-model, one-vote” SAT and PR distributions composed of the highest performing (closest to observed) member of each model ensemble (68 members total; see Fig. 1). Ensemble medians of the one-model, one-vote subsets are shown in solid black and listed in the upper right corner of panels c) and d). Panels e) and f) show where the 5 recommended independent high performers fall within the initial full CMIP5/6 JJA Central European mid-century change distributions.

Prior to computing the performance/independence cost function, we define a pool of high performers using a performance threshold of $D_i < 5$ (D_i ranges from 4.1 to 11.2). This eliminates lower performing (though possibly very independent) models from contention and reduces the selection pool from 68 to 20 models. All possible combinations of 5 out of the 20 models in contention are then iteratively evaluated for the magnitude of their performance/independence cost function, C_α defined as:

$$C_\alpha(m_1, \dots, m_5) = (1 - \alpha) \cdot \frac{1}{\min(D_i)} \sum_{i=1}^5 D_i - \alpha \cdot \frac{1}{2} \sum_{j \neq i}^5 \frac{S_{ij}}{\overline{S_{iN}}}$$

$$\overline{S_{iN}} = \frac{1}{N} \cdot \sum_{j \neq i}^N S_{ij}$$

In the first term, the performance portion of the cost function is the sum of D_i of the 5 chosen models divided by the minimum D_i of the highest performing model in the selection pool. In the second term, the independence portion of the cost function consists of the sum of the distances between a model and the other 4, S_{ij} , divided by the average distance of that model from all other models in the selection pool $\underline{S_{iN}}$. The factor of 1/2 in the independence term accounts for double counting of the distance between model i and model j and between model j and model i . A performance/independence trade-off factor, α , allows users to decide the importance of independence in model selection for their use case. Here, we choose $\alpha \approx 0.25$ to recommend AWI-CM-1-1-MR-r1i1p1f1, MRI-ESM2-0-r1i1p1f1, HadGEM3-GC31-MM-r4i1p1f3, NorESM1-ME-r1i1p1, and GISS-E2-1-G-r1i1p3f1 as five independent high performing simulations for Central European summer climate applications. For more information on model sub-selection as a function of α , please refer to Appendix A1.6.

MAIN RESULTS -- The ultimate goal of ClimWIP model sub-selection, shown in Fig. 2, is to provide a representative set of high performing, reasonably independent models for a task, in this instance, realistic mid-century Central European summer climate change. Therefore, we must evaluate where recommended independent high performers fall within the CMIP5/6 distributions of projected June-July-August (JJA) Central European SREX region (CEU; Iturbide et al., 2020) SAT (Fig. 2a,c,e) and PR (Fig. 2b,d,f) change between 1995-2014 and 2041-2060 average values. To contextualize the positions of the independent high performers, the CMIP5/6 JJA CEU distributions weighted by ClimWIP are shown in panels a and b for SAT and PR, respectively. As highlighted in previous WP2 reporting, ClimWIP provides a justifiable constraint on ensemble projection uncertainty. Here ClimWIP tends to give more weight to models that project around 2-2.5°C of JJA CEU warming by midcentury (in the RCP8.5/ssp585 scenario), with the exception of two high performing CMIP5 members with warming estimates at the upper and lower ends of the distribution (GFDL-CM3-r1i1p1 and GFDL-ESM2M-r1i1p1). ClimWIP also narrows the overall PR distribution towards a median PR reduction of -0.16 mm/day or an about 7% reduction from the observed 1995-2014 JJA CEU area average (not shown).

To reflect another common model sub-selection strategy, often referred to as the “one-model, one-vote” approach (Knutti et al. 2017), distributions that consist of the highest performing member from each model ensemble are shown in Figs. 2c and 2d. The distributions are broadly similar to their ClimWIP-weighted counterparts. For SAT, both the ClimWIP-weighted and one model, one vote distributions lack the clear bimodality of the full CMIP5/6 distribution (Fig. 2e, gray), but do not eliminate the possibility of 3-4°C midcentury JJA CEU SAT changes. The five independent high performers selected via the ClimWIP sub-selection protocol project midcentury JJA CEU SAT change along these lines (Fig. 2e, vertical lines); 4 of the 5 models project midcentury JJA CEU SAT changes of between 2°C and 3°C, around the ensemble median, while the remaining model, HadGEM3-GC31-MM, projects 3.5°C of warming by midcentury.

ClimWIP and the one model, one vote approach differ in how they constrain the high side tail of the midcentury JJA CEU PR distribution: ClimWIP more substantially reduces the likelihood of a future increase in JJA CEU PR. The difference arises in part due to the degrees of dependence between multi-model ensemble constituents; ClimWIP accounts for these degrees of dependence, while the one-model, one vote strategy assumes independence based on model name (Merrifield et al. 2020). Even so, a midcentury JJA CEU PR increase is unlikely in either case, a tendency also reflected by the 5 independent high performers (Fig. 2f, vertical lines). Ranging from -0.05 mm/day (2% reduction) to -

0.41 mm/day (17% reduction), the independent high performers span the constrained distributions with 2 models closely resembling the constrained ensemble median projection of approximately -0.15 mm/day and 1 model, again HadGEM3-GC31-MM, capturing the more substantial midcentury JJA CEU PR reduction of the ensemble's low side tail.

SUMMARY -- We demonstrate how ClimWIP, a multi-model ensemble weighting strategy developed as part of the EUCP's WP2, can be utilized to select independent, realistic model projections from the CMIP5 and CMIP6 archives for a given task. Here the ClimWIP sub-selection protocol is applied to the task of selecting 5 models suitable for the projection of midcentury summer Central European climate change under the RCP8.5/ssp585 forcing scenarios. The 5 recommended independent high performers are:

- AWI-CM-1-1-MR r1i1p1f1 CMIP6 historical + ssp585
- MRI-ESM2-0 r1i1p1f1 CMIP6 historical + ssp585
- HadGEM3-GC31-MM r4i1p1f3 CMIP6 historical + ssp585
- NorESM1-ME r1i1p1 CMIP5 historical + RCP8.5
- GISS-E2-1-G r1i1p3f1 CMIP6 historical + ssp585

based on their performance (with respect to 6 relevant observed climatological benchmarks) and independence (following Merrifield et al. 2020). The ClimWIP sub-selection protocol can be employed for any task for which identifiable performance metrics exist and can select subsets of any size; work is currently underway to provide 3, 5, and 10 model subsets for other European regions and seasons covered by the EUCP using predictors selected from a base set of climate fields (SAT, PR, SWCRE, SST and sea level pressure). In the near future, the ClimWIP sub-selection protocol will be integrated into the ClimWIP recipe on the ESMValTool for open use.

3.3 Science part 2 - CMIP6 sub-selection based on an assessment of key performance criteria (MO)

INTRODUCTION AND RATIONALE -- The aim in this study is to demonstrate a sub-selection of the CMIP6 ensemble for downscaling and regional impact assessment, where it may not be feasible to apply the full ensemble. Sub-selection might focus on selecting a subset that (a) samples from the range of future changes in relevant climate characteristics, and/or (b) excludes the least realistic models from which projections might be considered less reliable, or a combination of both of these criteria (e.g. McSweeney et al. 2015, Lutz et al. 2016).

Historically, climate modelling communities have been cautious about weighting or eliminating poor performing members due to the difficulties of linking baseline performance with projection plausibility defaulting to a 'one model, one vote' approach (e.g. Knutti et al. 2010, IPCC AR4 2013, IPCC AR5 2013). However, an increasing body of literature demonstrates cases where observed shortcomings in the ability of an individual model to realistically represent a baseline climatology are an indicator of reliability of that model's future projections (e.g. Whetton et al. 2007, Overland et al. 2011, Lutz et al. 2016, Ruane and Dermid. 2017). Further, for the purposes of applying models in impact studies at the regional scale, very unrealistic models may be of limited use if they require excessive bias correction. While these are strong arguments for filtering the ensembles for regional applications, the practical implementation requires us to navigate several challenges such as how to select appropriate criteria,

where the appropriate thresholds should lie for ‘acceptable’ versus ‘unacceptable’ models, and how to deal with models that perform well against some criteria but poorly against others.

In EUCP, we build on the approach developed in McSweeney et al. (2015, 2018), previously applied to CMIP5. In McSweeney et al. (2015, 2018), CMIP5 models were assessed on a range of regional criteria, including the circulation climatology, the distribution of daily storm track position, and the annual cycle of local precipitation and temperature in European sub-regions. These characteristics were assessed using a qualitative framework for flagging poorly performing models as ‘implausible’, ‘significantly biased’, or ‘biased’. This performance information was subsequently used together with information about projection spread (McSweeney et al. 2015) or model interdependencies (McSweeney et al. 2018) to arrive at subsets of the required size.

For CMIP6, we focus on building the regional performance ‘evidence base’ using a similar qualitative assessment framework as McSweeney et al. (2015). Here we classify models in a more informative way than simply keep or reject, to provide further information about how that model may perform in terms of key processes that influence the climate in a particular European region. These classifications can be broadly described as ‘implausible’ (unable to represent the present day climatology in a useful way), ‘unsatisfactory’ (containing substantial errors and relatively poor compared to other models) and ‘satisfactory’ (either lower than average errors or if more significant biases are present they are in limited in terms of the region affected).

The CMIP6 regional assessment summary will subsequently be combined with information about future projection range or model interdependencies (potentially drawing on the other methods of this report) to identify subsets of the ensemble. In this science part, we will demonstrate the application of the performance information to identify CMIP6 subsets of a number of sizes (for example, $n=5$, $n=9$, $n=15$) which will be suitable for a range of applications.

PERFORMANCE ASSESSMENT CRITERIA -- Classification definitions -- We have chosen a traffic light style approach to model classification; we initially start by ruling out models that are unable to represent the criteria in any useful way. **Red** is used for this category and is defined as: unable to represent the present-day climatology in a useful way. Red is used to flag models that are considered implausible and should be ruled out from sub-sampling on the basis of a single criteria. The other classifications used are as follows:

- **Orange** Substantial errors and bias in remote regions and present in the local region of interest.
- **Yellow** Some substantial errors present, but not widespread or not present in the local region of interest.
- **White** Lower than average error over both local and remote regions.
- **Grey** Data/analysis not available.

The selection of the individual criteria and examples of how the assessment has been carried out is described in the following sections. The application of this assessment to sub-selection is then described in the last section.

PERFORMANCE ASSESSMENT CRITERIA -- A number of criteria are selected based on their importance for the European climate, key variables such as temperature and precipitation, along with others that

indicate how well models are able to represent key processes that drive the present day and future climate changes for Europe.

The near surface temperature and precipitation are key variables for future climate and are of primary consideration in impact studies, especially in terms of future hydrology considerations (e.g. White et al 2011, Ruane et al. 2014, McDermid et al 2015). They have been considered key variables in previous sub-setting approaches (e.g. Ruane and Dermind 2017, McSweeney et al. 2015).

A number of previous studies have considered the importance of capturing the main synoptic features and large-scale atmospheric circulation patterns (e.g. McSweeney et al. 2015, Prein et al 2019) as a key criteria for GCM sub-setting. For northern Europe in particular large scale weather patterns and the prevailing South Westerlies are essential to representing the climate. In the winter (DJF) many models have a southern bias in the flow and the winds become too zonal with a positive bias in the windspeed across central Europe. In the summer (JJA) the prevailing Westerlies are less strong but still an important driver of weather systems and key for representing the climate. We assess the models by comparing a 20 year baseline climatology with the ERA data, using a similar approach to McSweeney et al. (2015).

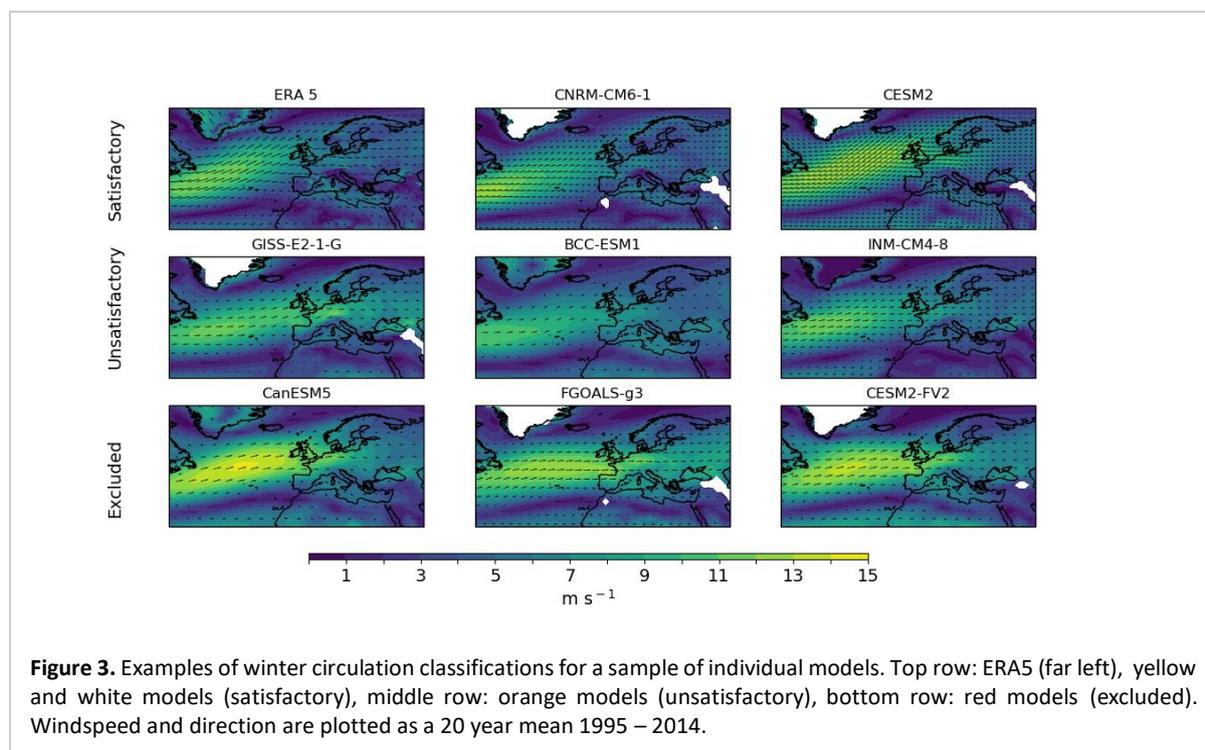
These features along with the North Atlantic storm track are important for representing the climate and the potential future impacts of changes in the climate. Extratropical cyclones are the dominant weather type in mid-latitudes where they can have a significant impact due to extreme precipitation and windspeeds (Browning 2004, Priestly et al. 2020). They have an important role in the general circulation in the poleward transport of heat, moisture and momentum (Kaspi and Schneider 2013) and in maintaining the latitude westerly flow. We use the analysis of Priestly et al. (2020) to assess the individual CMIP6 models for the main North Atlantic region and the individual European regions. Blocking by high pressure weather systems is known to cause periods of cold dry weather in the winter and summer heat waves. Blocking is typically underrepresented in GCMs and this is still the case in large parts of Europe in CMIP6, although there has been some improvement in the bias in many CMIP6 models (Davini and D'Andrea 2020). We use the results of the analysis carried out by Davini and D'Andrea (2020) to assess the performance of the CMIP6 models based on RMSE, bias and correlation.

Previous studies have indicated that North Atlantic Sea Surface Temperature (NASST), including the Subpolar Gyre (SPG) are a significant driver of variability in the European climate (Friederichs and Hence 2003, Sutton and Hodson 2003, Sutton and Buwen 2012, Dong et al 2013, Menary et al.,2021). The AMOC also plays a significant role in determining the present and future NASST, along with the European climate due to its role in the poleward transfer of heat and ocean circulation. The CMIP5 and CMIP6 ensemble both predict a reduction in the AMOC by the end of century for higher emissions pathways and it is likely that changes in the AMOC and NASSTs may play an increasingly important role in future changes to the temperature and precipitation patterns in Europe. We use comparisons of spatial maps of SST temperature bias in the NA along with comparisons of the SPG gradients with reanalysis data to classify SST performance. The AMOC model comparison with rapid data from the analysis of Menary et al. (2020) is used to assess the AMOC in each model.

PERFORMANCE ASSESSMENT EXAMPLES -- Examples of how the classifications have been applied to the models for winter circulation are shown in Fig. 3. The top 2 models (to the right of ERA5) have

been taken from the yellow and white classification and are considered ‘satisfactory’ models. In comparison to the ERA5 data they have well represented wind patterns, and magnitude biases in the windspeeds are relatively small. The unsatisfactory models have a wind pattern that is too westerly, and windspeeds that are too strong over central Europe. The excluded models have an almost entirely westerly (no south-westerlies) wind direction and large positive bias over Europe.

Comparison of the models with the observed precipitation cycle for central Europe is shown in Fig. 4. An example assessment table (for a selection of models) for the large-scale criteria, which is applied for all regions is shown in Table 1.



USING PERFORMANCE INFORMATION TO IDENTIFY SUITABLE SUB_SETS -- We aim to use the results of the assessment described in the previous section to identify a suitable sub-set from the CMIP6 ensemble for each European SREX region. An example of the assessment information for the criteria that will be used in this method is shown in Table 1. If a model has been flagged red for one of the assessment criteria, it is removed from the subset as this indicates that it is unable to represent a key process driving the present-day climate and it is reasonable to assume that changes in the climate projected by the model cannot be interpreted in a meaningful way. The intention is to produce a subset of models that captures the full ensemble range that includes all models that can be considered to provide plausible projections. We also aim to further select for models that are considered to perform better against the assessment criteria, while maintaining this projection range, see Fig. 5 for an overview of the method. The impact on the projections for each of the SREX regions of applying these criteria as outlined in Fig. 5 to the temperature and precipitation projection spread (for ssp585, end of century) is shown in Fig. 6. It is noted that this approach to sub-selection eliminates more models that show relatively modest changes, than those with higher climate sensitivity.

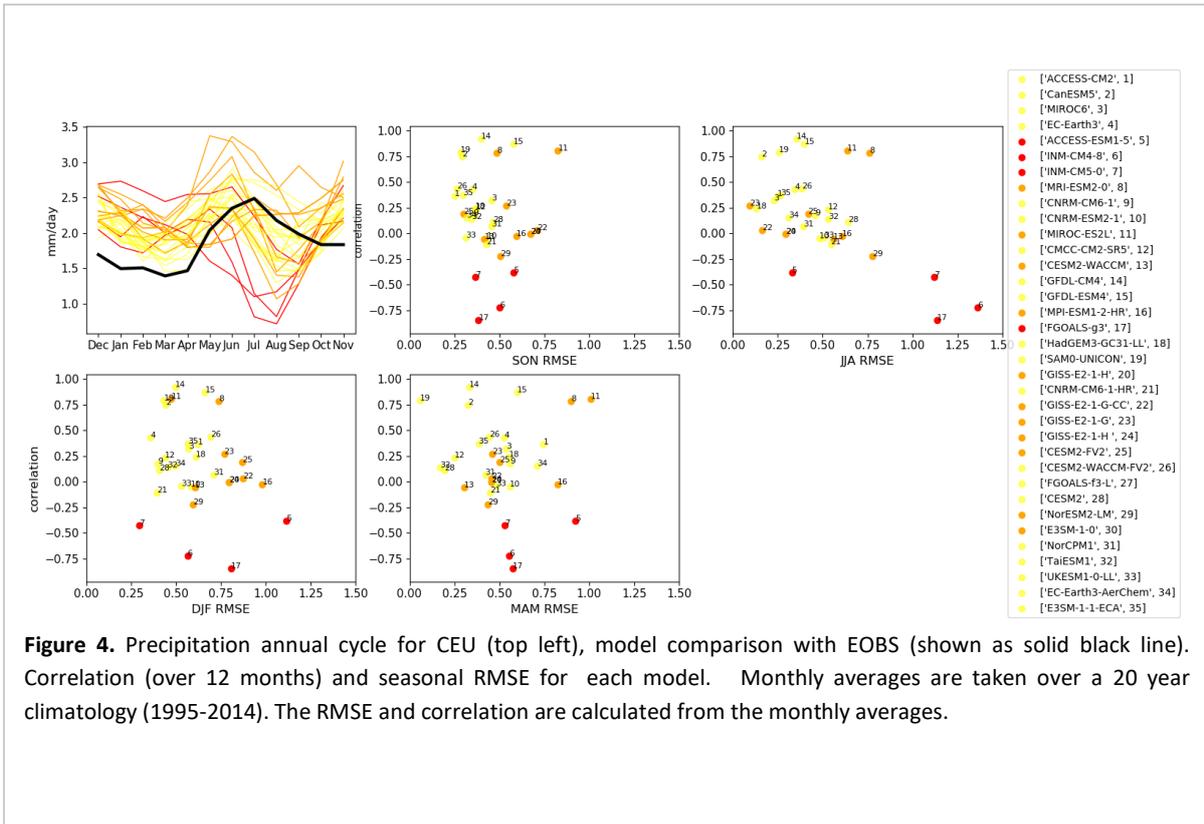


Figure 4. Precipitation annual cycle for CEU (top left), model comparison with EObs (shown as solid black line). Correlation (over 12 months) and seasonal RMSE for each model. Monthly averages are taken over a 20 year climatology (1995-2014). The RMSE and correlation are calculated from the monthly averages.

NEXT STEPS -- The aim of this approach is to produce recommended subsets for each of the European SREX regions. These subsets will be of a range of sizes from about 15 models (as a subset of plausible GCMs), to 5 models (for downscaling, this subset will also be dependent on data availability). For further thinning of the ensemble, such as in the case of downscaling a number of options are available to the end user. From the assessment information available preferred models that are classified as satisfactory against the criteria can be selected where this allows for the ensemble range to be maintained. An alternative approach could be to thin the remaining ensemble by removing nearest neighbours (where models share large amounts of the same code).

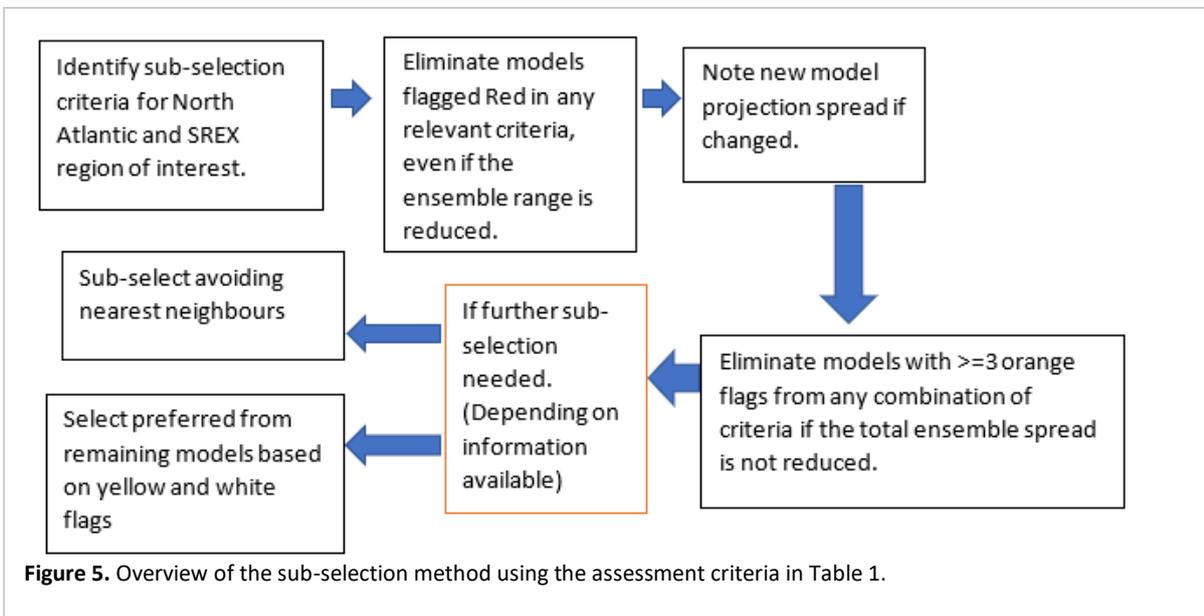
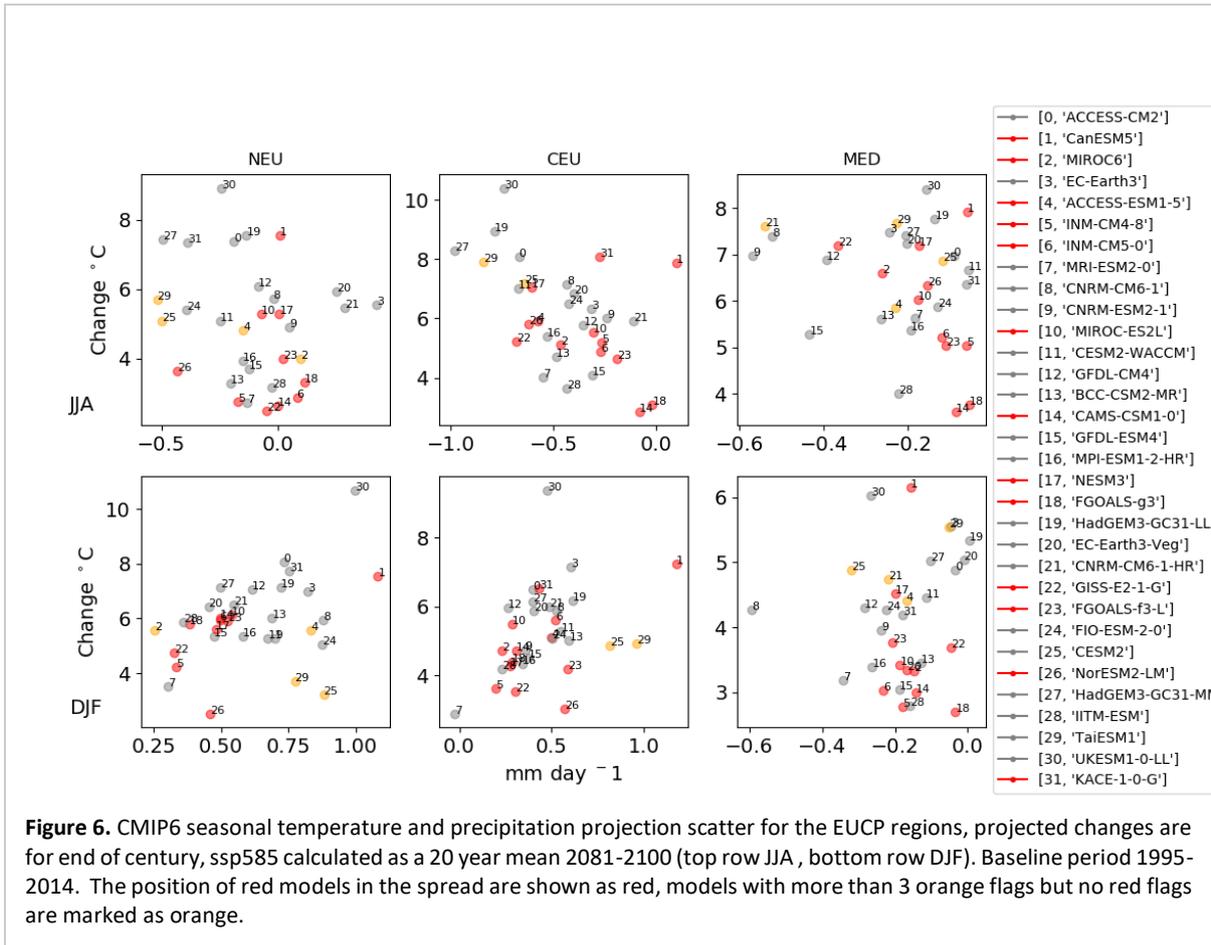


Figure 5. Overview of the sub-selection method using the assessment criteria in Table 1.

Table 1. Sample assessment table for large scale criteria for a selection of CMIP6 models. A sample of models are shown here, more model assessments data are available in Appendix A2 along with the flags for each of the SREX regions.

All regions									
	Seasonal mean blocking frequency (DJF)	Mean DJF circulation	Seasonal mean blocking JJA	Mean JJA circulation	Atlantic SST DJF	Atlantic SST JJA	Storm Track DJF	Storm Track JJA	A M OC
AWI-CM-1-1-MR									
AWI-CM-1-1-LR									
ACCESS-CM2									
ACCESS-ESM1-5									
BCC-CM2-MR									
BCC-ESM1									
CAMS-CSM1-0									
CESM2-WACCM-FV2									
CESM2-WACCM									
CESM2									
CESM2 – FV2									
CMCC-CM2-SR5									
CNRM-CM6-1-HR									
CNRM-CM6-1									
CIESM									
CNRM-ESM2-1									
CanESM5									
EC-Earth3-Veg									
EC-Earth3									
FGOALS-g3									



3.4 Science part 3: Towards subselection based on circulation changes: changes in Jet Latitude and Strength and the Summer NAO (MO)

The UK Met Office has been exploring ways to characterise projection uncertainties according to large scale physical changes. By understanding how changes in key circulation characteristics of the regional climate drive local characteristics of climate change (e.g. mean precipitation changes), we can potentially identify useful axes against which to sub-select and develop storylines of future change that can be used to communicate coherent scenarios of projected changes.

We have explored a number of different ways of characterising large-scale characteristics of European climate. These approaches include a number of circulation indices such as weather types based on a clustering approach, blocking indices, storm tracks and the jet stream. Our analysis is based on the Met Office’s 15-member PPE (‘PPE-15’) (Murphy et al., 2018) and a 13-member subset of high-performing members of CMIP5 (McSweeney et al., 2018), which comprise the UKCP18 global dataset.

JET STREAM STRENGTH AND LATITUDE -- Here we show an application of an index describing the latitude and strength of the Eddy-driven jet stream over the north Atlantic. The diagnostics of jet latitude and strength we use here are based on those of Woollings et al. (2010). The latitudinal positions of the jet stream are characterised in autumn and winter by a distinctive trimodal

distribution, with ‘preferred’ positions at around 35°N, 45°N, and 55-60°N, and tending towards the two northern positions and weaker flow in summer and spring (Fig. 7, Top row).

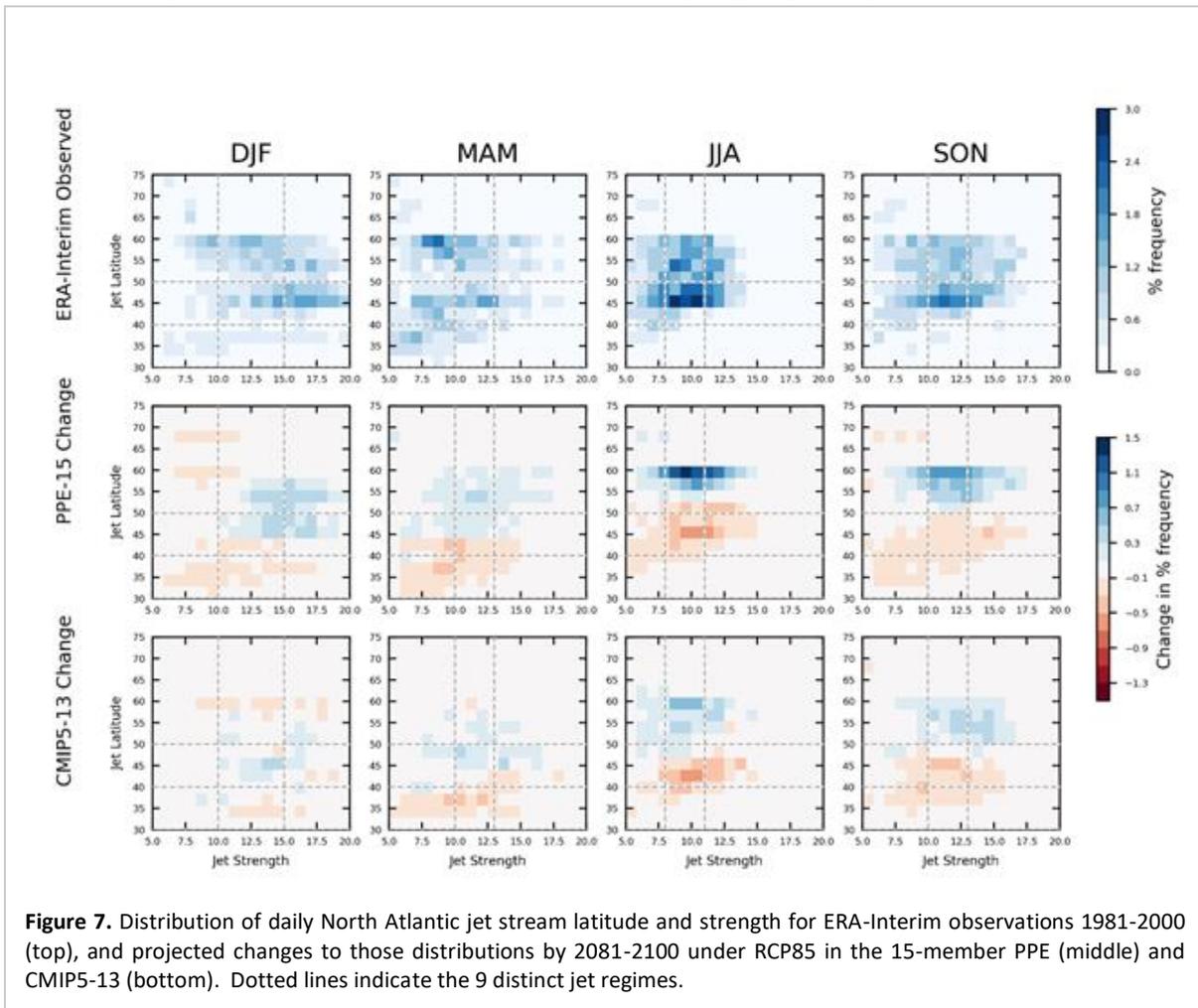


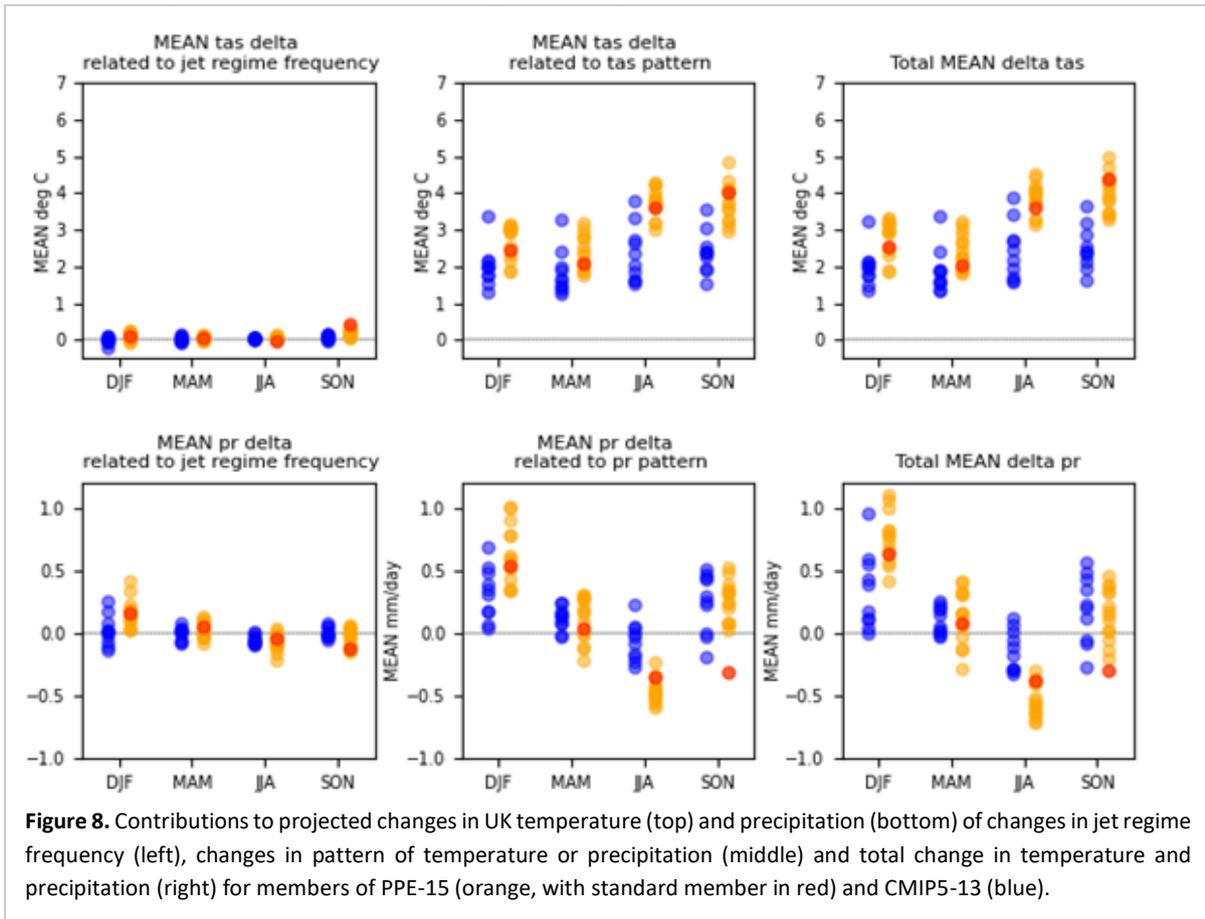
Figure 7. Distribution of daily North Atlantic jet stream latitude and strength for ERA-Interim observations 1981-2000 (top), and projected changes to those distributions by 2081-2100 under RCP85 in the 15-member PPE (middle) and CMIP5-13 (bottom). Dotted lines indicate the 9 distinct jet regimes.

Here we link projected changes in the jet stream to changes in regional temperature and precipitation. Fig. 7 (middle and bottom rows) shows that projections from PPE-15 and CMIP5-13 indicate a general tendency for a northwards shift of the jet. In the PPE-15, this manifests as an increase in days where the jet occupies in central and northern positions, as well as greater strengths in SON and MAM, while for CMIP5-15 this signal is less clear. Both ensembles indicate for JJA and SON a shift towards high frequency of days when the jet occupies the northern position, and fewer days in the southern and central positions.

We explore changes in projected mean temperature and precipitation by assigning each day to one of 9 ‘jet regimes’ based on jet strength and latitude, as indicated by the dotted lines in Fig. 7. These 9 regimes result in distinct geographical distributions of precipitation and temperature, particularly when the jet is strong. Changes in mean temperature and precipitation can subsequently be disaggregated into those associated with the change in jet regime frequency (dynamical changes) vs those resulting from a change in the rainfall or temperature pattern associated with that jet regime.

This is illustrated for the PPE-15 ensemble mean in Fig. 8, in which the pattern of mean rainfall change associated directly with the changes in jet regimes can be seen in the left hand column. The increase

in days in autumn/winter where the jet is strong/north at the expense of days south/weaker jets lead to increases in mean rainfall in Northern Europe, particularly the UK, and decreases in mean rainfall in southern Spain. Decreases in rainfall in southern Europe as a result of fewer days when the jet sits in the southern positions are evident throughout the year.



The pattern of mean rainfall change that is not associated with the jet regime (the thermodynamic component), are shown on the right in Fig. 9. These changes remain larger than the dynamical component, and as expected represent a pattern of summer drying and winter rainfall increases that might be expected as a result of a warmer atmosphere. However, we should expect that this ‘residual’ change also still includes some degree of dynamical changes not captured by the 9 jet regimes.

The ensemble mean in Fig. 9 hides the regional variations across the ensemble which are of interest for identifying storylines which capture model uncertainty. The ensemble spread in these quantities is summarized for the UK as an example region in Fig CM3. For the UK, the changes in distributions of jet regimes contributes to the spread in DJF rainfall changes.

Further work on these jet regimes will explore suitable metrics for selecting members that encapsulate a wide range of plausible changes with direct relevance to Europe, for example, adding a dimension to the criteria for sampling for approaches such as the sub-selection in section 3.3 of this report). The development of the jet latitude and strength ‘regimes’ and their comparison with historical simulations is being prepared for publication (McSweeney et al, in prep).

EXPLORING CHANGES IN THE SUMMER NAO AS A DRIVER OF REGIONAL PRECIPITATION CHANGES -

- The SNAO is defined as the leading EOF of summer SLP variability in the North Atlantic/European Sector, which has been shown to modulate precipitation in Northwest Europe (including the UK) and the Mediterranean. More specifically, a positive SNAO index represents high average summer pressure anomalies over the UK, consistent with lower-than-average seasonal rainfall. Blade et al. (2012) demonstrated that a projected upward trend in the SNAO in CMIP3 models can be considered to be a driver of a consistent signal of precipitation decreases in northwest Europe, and also a source of uncertainty given that the magnitude of SNAO change, and the magnitudes of summer rainfall decreases vary significantly across members. Following the methodology of Blade et al. (2012), we have assessed the extent to which the PPE and CMIP5-13 models represent the SNAO and its teleconnection to European summer precipitation, and explored the relationship between the magnitude of projected changes in SNAO and UK summer rainfall.

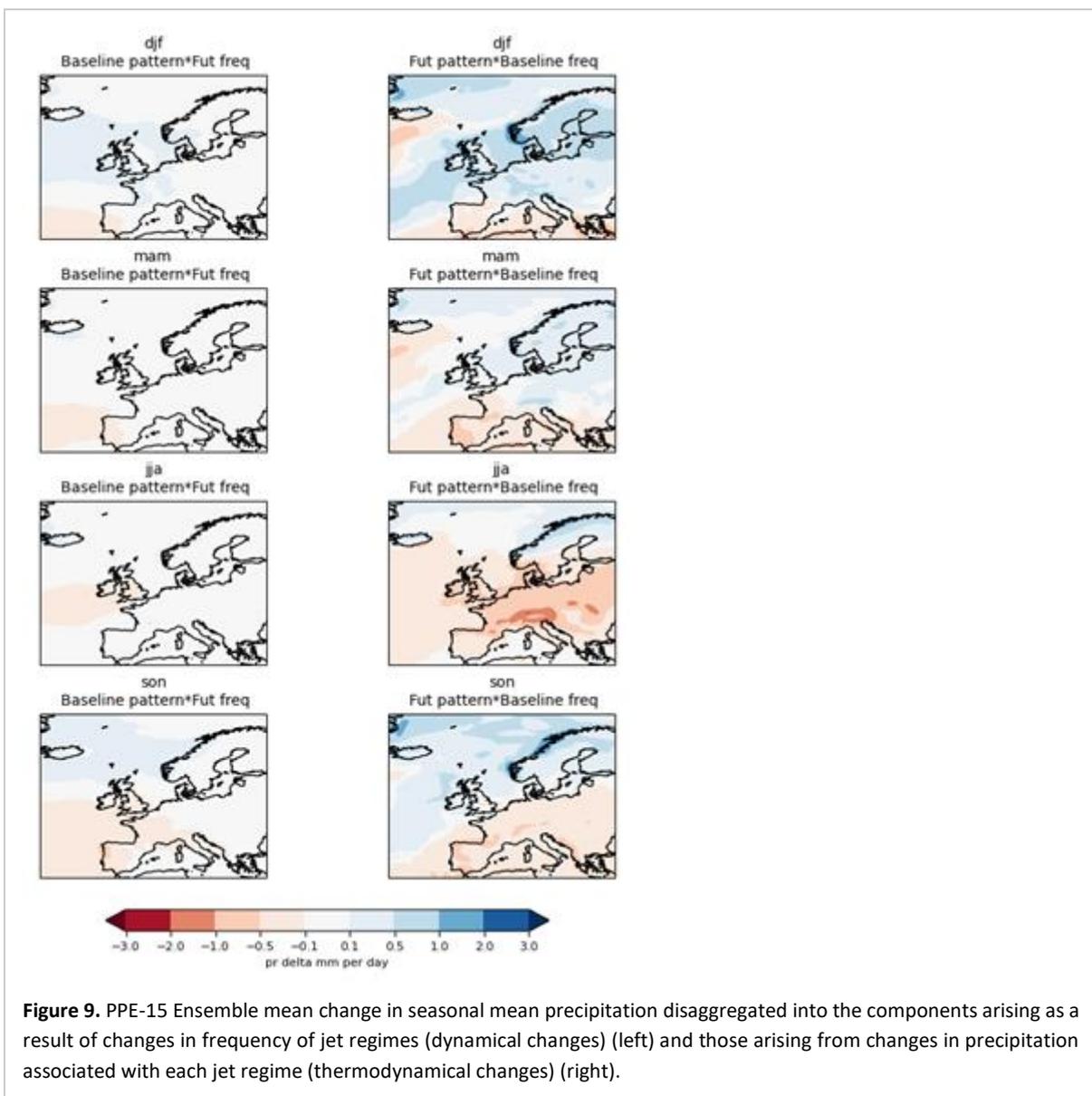
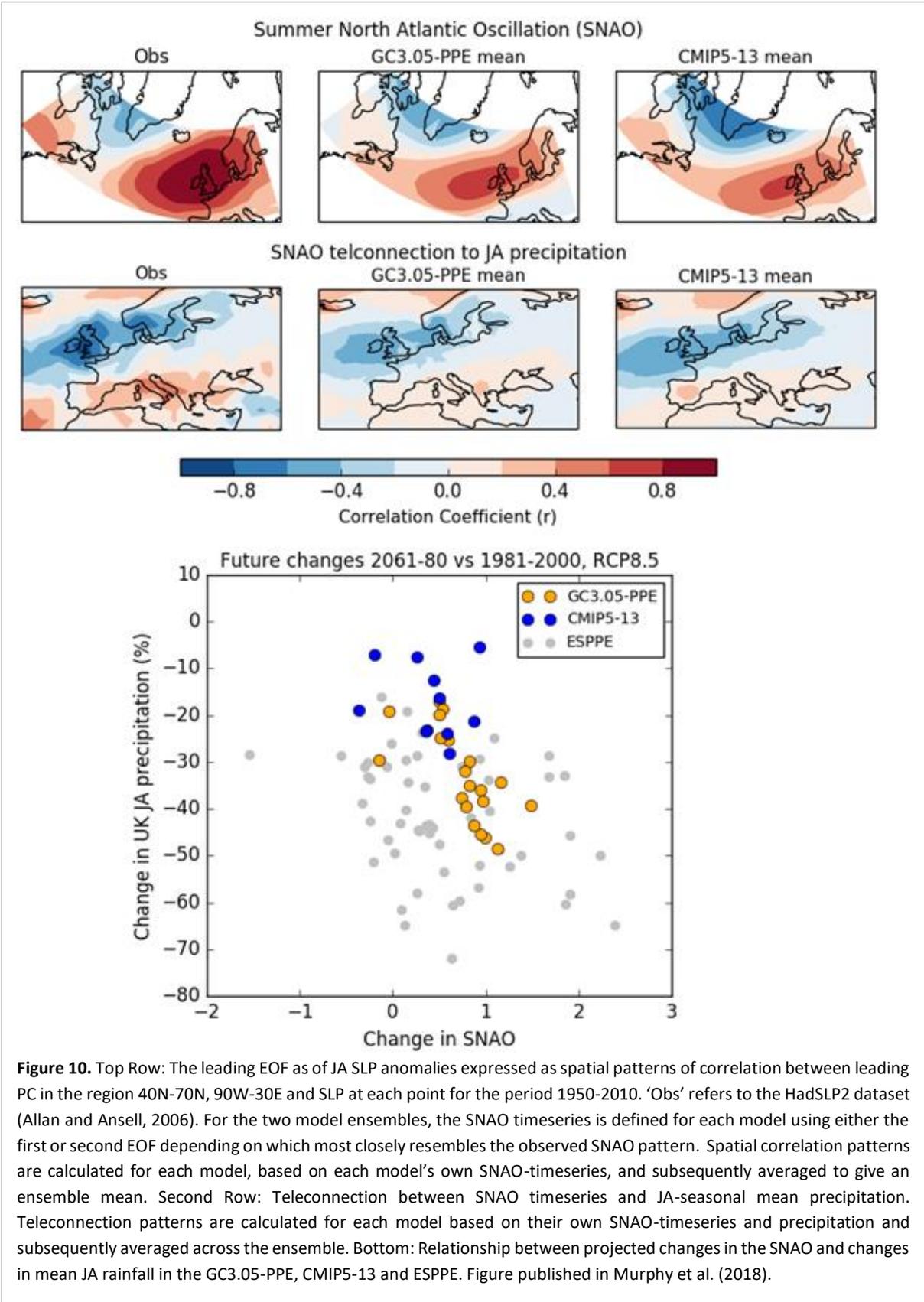


Figure 9. PPE-15 Ensemble mean change in seasonal mean precipitation disaggregated into the components arising as a result of changes in frequency of jet regimes (dynamical changes) (left) and those arising from changes in precipitation associated with each jet regime (thermodynamical changes) (right).

Most models in both PPE and CMIP5-13 simulate a spatial SLP pattern with SNAO-like north-south dipole (ensemble mean SLP patterns are shown in Fig. 10 top row, individual models are not shown), and the teleconnection patterns shown in Fig. 10 second row indicate that the ensembles broadly capture the negative correlation between SNAO and July-August (JA) precipitation over the UK and NW Europe (although this is not the case for all the members with the ensembles, several less-realistic members within both ensembles position the southern dipole too far west into the Atlantic, and or fail to represent the teleconnection with UK rainfall). However, what is also noted in Blade et al. (2012), and is also demonstrated here for the PPE is the weaker than observed positive correlation relationship between Mediterranean rainfall and the SNAO, suggesting that models are missing or underestimating processes which link the large-scale circulation with Mediterranean rainfall.

Given that the models capture both broadly realistic SNAO-like variability and this has been shown to explain a significant proportion of the variability in UK JA precipitation, we might expect to see similar relationship between changes in the SNAO and JA precipitation in future projections, as reported by Blade et al. (2012). Indeed, Fig. 10 bottom, indicates that most members of the three ensembles project a positive shift in the SNAO, and that those with larger increases in SNAO tend to display larger reductions in JA rainfall, although there remains significant scatter in the relationship. Notably, the larger increase in SNAO may partially explain the greater summer drying in PPE relative to CMIP5-13.

The SNAO therefore offers useful 'axis' along which to sample from global models ensemble in order to capture spread in European rainfall, and we note that capturing the range of plausible changes in the SNAO may require us to look beyond the CMIP ensembles - e.g. by drawing also on the UKCP PPE in order to include members with the largest SNAO changes and related precipitation changes. Further exploration of the relationship between the SNAO and Mediterranean rainfall clearly warrants further investigation in order to understand the missing teleconnection and its implications for projections of future mediterranean rainfall.



3.5 Science part 4 - Selecting regional simulations tailored to user's needs and setting small (CPRCMs) and large (GCMs) ensembles into uncertainty context (SMHI)

INTRODUCTION AND RATIONALE -- As climate change awareness is rapidly increasing and adaptation and mitigation measures are developed and taken into action, stakeholders are in need of holistic climate information on regional and local scales. With holistic information we mean to transfer to stakeholders a deeper understanding of climate simulations which goes beyond single numbers of raising temperature levels. In this way the stakeholder community learns more about what climate simulations are, not perfect forecasts but rather a range of possible futures. For many years the climate community has been producing more and more climate simulation and data. From being overwhelmed the climate impact community grew into better understanding of the impact/meaning of climate simulations and the multitude of models. Still the sheer amount of available simulations to choose from has been a hurdle in discussions with decision makers. To guide users of climate information through the increasing amount of data, a model selection process was developed to condense the ensemble size of regional climate models. The process uses two moments: first the focus and interest of the climate data user is carved out and then a smaller subset of the ensemble calculated with the aim to keep the variability within user relevant climate information (Wilcke and Barring, 2016) -- Providing information about the uncertainty of climate projections is deemed essential in communication with stakeholders. This requires large ensembles of climate simulations. However, CPRCMs are computationally too expensive and time consuming to downscale larger ensembles of GCMs and RCMs. Furthermore, in many projects and experiments the selection of which GCMs to downscale occurs based on various reasons, like financial, political, but also practical reasons, e.g. data availability as not all GCMs provide data on model levels which is needed for downscaling. Therefore in order to evaluate the background uncertainty, we set the CPRCM simulations produced within EUCP WP3 into the context of the RCM and GCM ensembles downscaled by the CPRCMs, and the other available RCM ensembles from EURO-CORDEX and PRINCIPLES (meaning all on esgf available RCM data on EUR-11 domain) and GCM ensemble of CMIP5. The EURO-CORDEX ensemble as well as the CMIP5 ensemble and subsets of those are widely used by climate impact researchers and stakeholders.

SELECTING REGIONAL SIMULATIONS TAILORED TO USER NEEDS -- The first moment is a dialog with the data users to provide understanding on both ends of what is needed (climate indices relevant for the user) and what is possible to deliver (a number of climate simulation output data that is still descriptive of the whole ensemble of simulations). This dialog and building of understanding is essential for a user tailored product. That means agreeing on e.g. seasons and regions which are focussed on in the study, climate variables like precipitation and indices like number of frost days, as well as distilling out if climate change signals are relevant or rather period averages and for which periods.

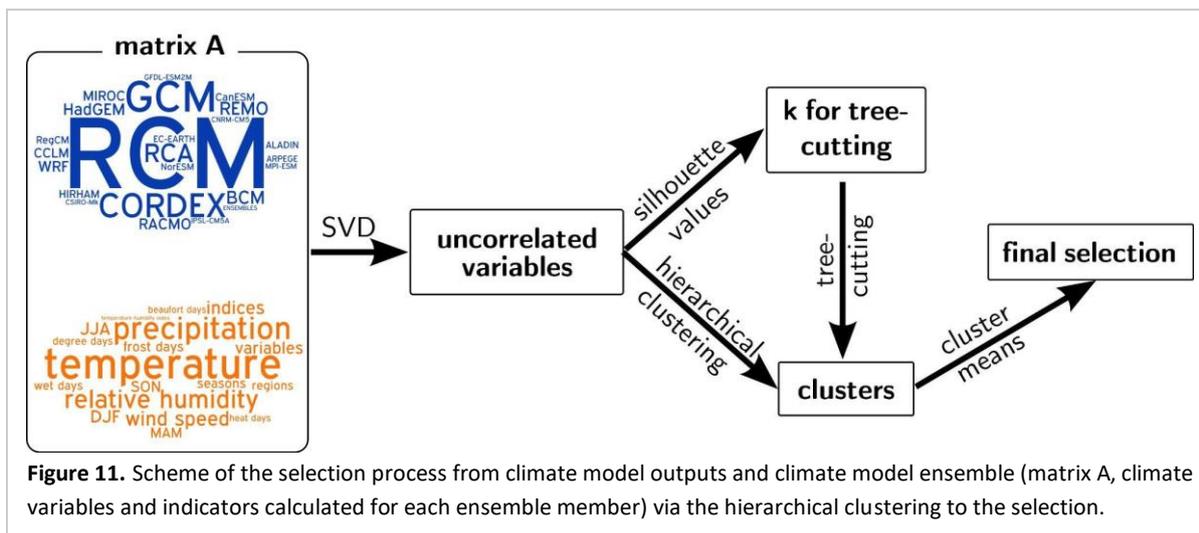


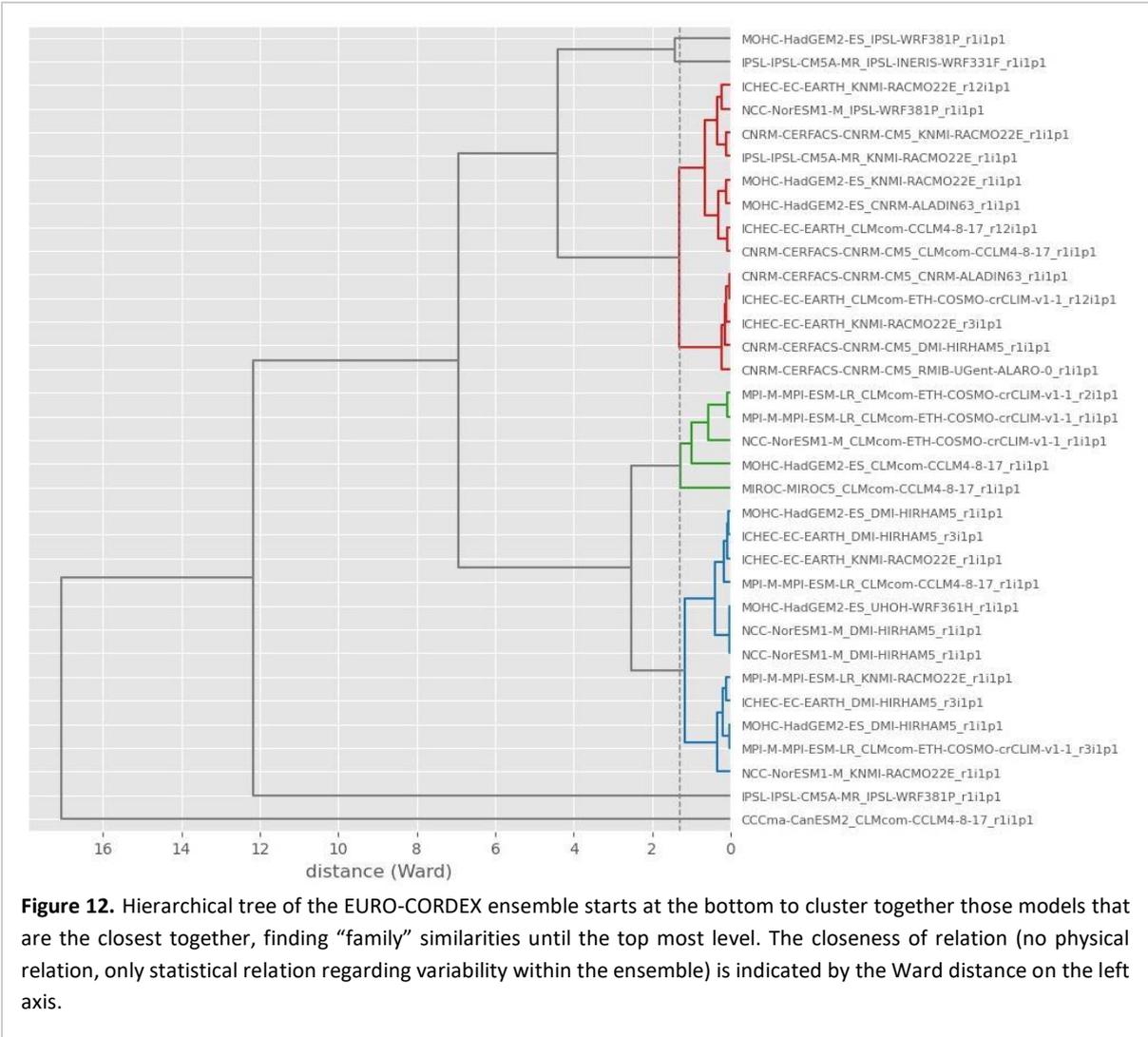
Figure 11. Scheme of the selection process from climate model outputs and climate model ensemble (matrix A, climate variables and indicators calculated for each ensemble member) via the hierarchical clustering to the selection.

The second moment is to calculate those indices for each ensemble member and transform the resulting matrix into an uncorrelated variable space using singular vector decomposition. This matrix is the information which we base hierarchical clustering on. Like the methods described before, we try to group similar behaving simulations together to then select just one out of each group. To support the decision on the optimal number of clusters, we introduced a step to calculate mean silhouette values for each number of cluster possibilities and chose the number of clusters/subsets with the largest distance to each other.

CASE STUDY 1 -- In a case study analysed together with EUCP WP4 we did a clustering for hydrological modelling. Here we followed the user's suggestion of indices which were all based on daily precipitation values. They covered annual and seasonal climate change signals (mid century) averaged for three regions in Europe. Fig. 12 shows the hierarchical clustering tree for 32 regional simulations.

The suggested selected simulations are:

- NCC-NorESM1-M downscaled by CLMcom-ETH-COSMO-crCLIM-v1-1
- MOHC-HadGEM2-ES downscaled by CNRM-ALADIN63
- ICHEC-EC-EARTH downscaled by KNMI-RACMO22E
- CCCma-CanESM2 downscaled by CLMcom-CCLM4-8-17
- MOHC-HadGEM2-ES downscaled by IPSL-WRF381P
- IPSL-IPSL-CM5A-MR downscaled by IPSL-WRF381P
- IPSL-IPSL-CM5A-MR downscaled by IPSL-INERIS-WRF331F



CASE STUDY 2 -- In a second case study the model selection was part of building a decision basis for downscaling a subset of a large single model ensemble (S-LENS, EC-EARTH) with an expensive convection permitting model. Here we framed the idea to pre-select members which trigger possible different types of convective precipitation events. The background to that was the user’s need to reduce the ensemble for being able to save only relevant model levels for further downscaling. The indices were based on climate change signals of convective precipitation and temperature for summer and winter seasons to mid century with focus on regions in Sweden. The clustering resulted in 12 distinct clusters and therefore 12 selected ensemble members which will add to an informed decision on which members to save model levels for and possibly downscale on very high resolution.

SETTING HIGH-RESOLUTION CPRCMs IN THE CONTEXT OF THEIR DRIVING RCMS AND GCMS -- The first step is to see where in the RCM and GCM variability space the chosen CPRCM simulations are placed. This is done for the Alpine domain as defined by WP3. The Alpine domain is the common region for which all institutes had to contribute. Therefore the CPRCM ensemble is largest for this region. The intention is to visualize how well the distributions of GCMs and RCMS are sampled.

Figure RW3 shows the change signals for 99th percentile of summer (JJA) precipitation (y-axis) against temperature (x-axis) for the Alpine region as defined by WP3. The change signals were calculated for the CMIP5 ensemble (orange dots, rcp 8.5) plus an EC-EARTH member from KNMI that the KNMI-CPRCM simulation downscaled (Appendix Table A3a). For the RCM ensemble on EUR-11 domain (blue triangles) we used simulations produced by European projects (EURO-CORDEX, PRINCIPLES, and REKLIES), plus the intermediate RCMs which were downscaled by the CPRCMs, that are not necessarily on EUR-11 but on similar domains and resolutions (Appendix Table A3b). The change signals of CPRCM simulations from WP3 (Appendix Table A3c) are shown as green crosses. The change signal is calculated from two periods available for the CPRCMs, the historical period 1996 - 2005 and the scenario (rcp8.5) period 2041 - 2050. The variability in the GCM (79 members) and RCM (64 members) ensembles is comparable. The CPRCM ensemble captures most parts of the temperature range but only the middle part of the precipitation distribution, remembering that this is the 99th percentile of the summer precipitation and temperature, respectively. Beside the key-variables (temperature and precipitation) the intention is to evaluate more climate variables and in particular indicators which are relevant for climate impact applications. The intention is to provide a ground and more comprehensive picture to users of very high resolution climate data.

In the second step the CPRCM simulations are connected to their driving RCM and GCM simulations. In this way the impact of the different downscaling steps on climate change signals becomes visible. For this step it is desirable to have an as large as possible ensemble of CPRCM simulations. WP3 used a common domain over the larger Alpine region that all institutes have to contribute to with their simulations. This Alpine domain is the initial focus of the model chain study. Fig. 14 shows the 99th percentile of summer temperature and precipitation change for the Alpine region and shows the same values as Fig. 13, however, only for the models that are part of the CPRCM downscaling chain (see Appendix Table A3c). One intermediate simulation (RCM) is missing only due to lack of accessibility to the data. Still the CPRCM (ETHZ-2) and GCM (MPI-ESM-LR_r1i1p1) is included in the figure to show the effect of the downscaling even with less details. For the 99th percentile of summer precipitation and temperature change no common pattern in how regional models affect the GCM signals becomes visible. The same holds true for the average change (not shown here). However, from this figure a certain narrowing of the precipitation distribution becomes evident. This is interesting because precipitation is naturally the focus variable when studying convective permitting climate simulation data.

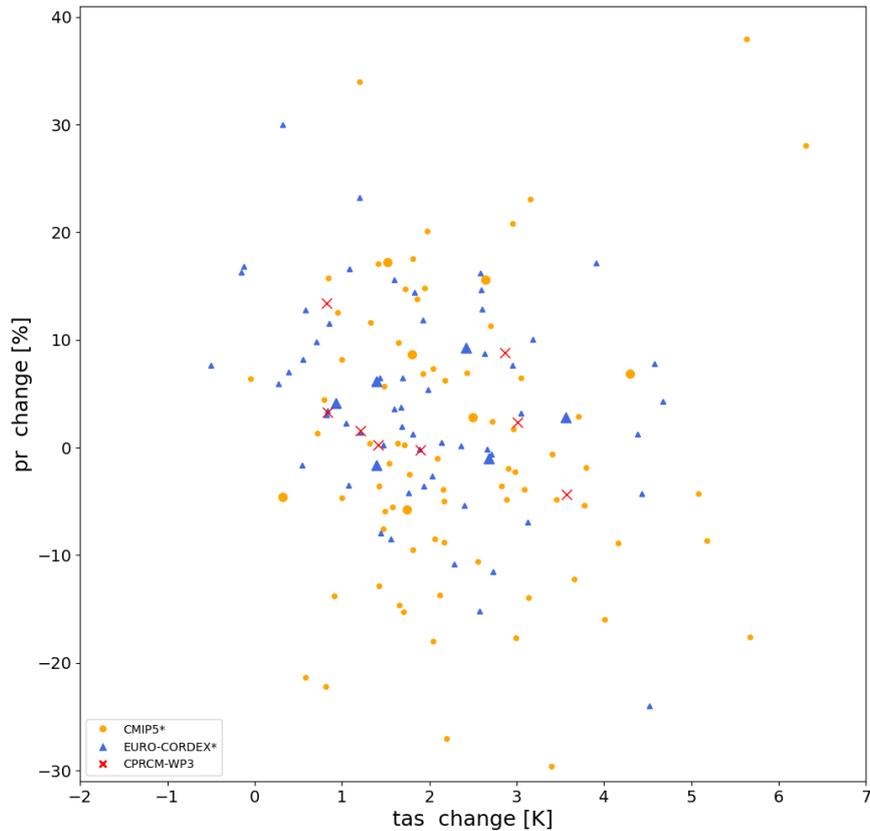


Figure 13. Climate change signal (2041-2050 minus 1996-2005) for the Alpine region (defined by WP3) for 99th percentile for summer (JJA) precipitation and temperature. Calculated for CMIP5 ensemble including KNMI-EC-EARTH-r14i1p1 (orange dots), EUR-11 ensemble (called EURO-CORDEX* as extended by other projects; blue triangles) including the RCMs that were downscaled by the CPRCMs but are not available on ESGF, and the CPRCMs from WP3 (red crosses). The RCMs and GCMs belonging to the downscaling chain of CPRCMs are enlarged.

In addition to the Alpine region, each WP3 modelling institute agreed to simulate over at least one more domain in Europe. Due to computational, financial and temporal limits those non-Alpine domains are covered by CPRCM ensembles with sizes ranging between 2 and 4 members. By studying the statistical changes in the variability space introduced by the downscaling steps a possible extrapolation of the variability can be made towards non-Alpine domains. The knowledge about the ensemble variability and the impact of downscaling on that variability, gained from the study over the Alpine region, lays the ground for improving uncertainty information in regions with very small ensembles or only single downscaling simulations.

CASE STUDY - GCMs - Large ensembles -- Another attempt to analyse uncertainty in an applied study was carried out for the summer 2018 heatwave in northern Europe. Two long-lasting high-pressure systems led to an extended and very warm summer season over Scandinavia. To assess the chances of such a heatwave happening, and find a possible attribution to climate change, we used 4 large single-model GCM ensembles and one multi-model ensemble (CMIP5). The large ensembles provide an enhanced sample size of 20580 summers representing summer over the last 70 years. The uncertainty or spread of the model ensembles enabled us to narrow down information about extreme events. By analysing the probabilistic distribution functions of monthly and seasonal (May to August) mean temperatures (Fig. 15) as well as some heat indices (not shown), it became clear that the observed 2018 summer (in particular the months of May and July) had been extremely warm. The

large ensembles indicate that such high temperatures could have happened in pre-industrial times, but at a lower frequency of occurrence. We can clearly see an increase in probability for such warm summers over Scandinavia even from the recent past (1951-1980) to present day climate (1989-2018) (Wilcke et al. 2020).

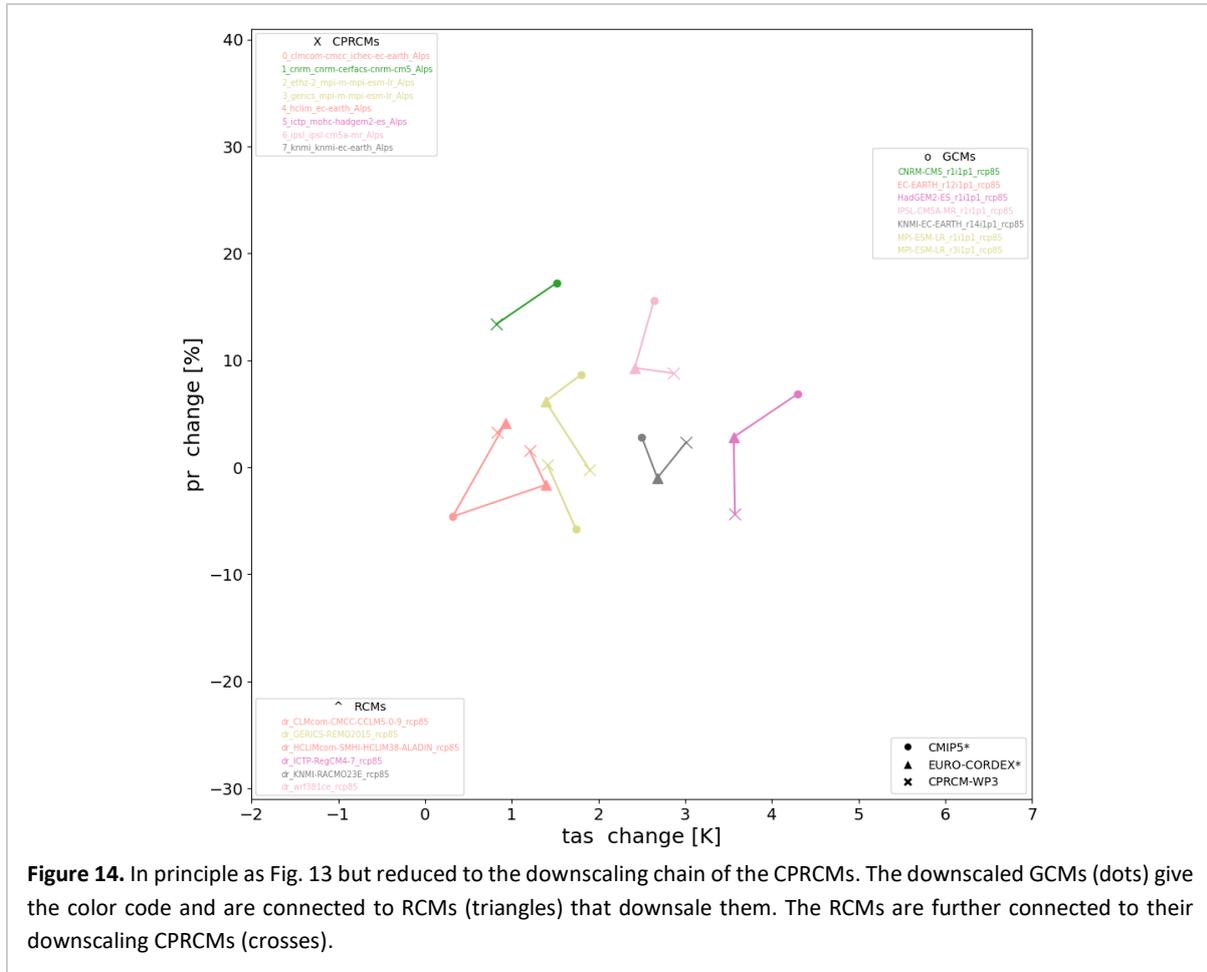


Figure 14. In principle as Fig. 13 but reduced to the downscaling chain of the CPRCMs. The downscaled GCMs (dots) give the color code and are connected to RCMs (triangles) that downscale them. The RCMs are further connected to their downscaling CPRCMs (crosses).

Wrap up -- In the first study climate scientists and climate data users engaged form detailed information for applying user tailored model selection. This is a way to reduce the ensemble size focussing only on specific interests of a user, meaning climate indicator. The selection tool shows a possibility to provide more user-friendly climate data without missing out on the uncertainty aspect of climate modelling. As in practice, the selection process is not always feasible or applicable, a “hindcast” analysis was carried out to set a selected subset of GCMs and downscaling CPRCMs into the context of the larger GCM ensemble. The case is taken from WP3 where the selection of GCMs was guided by CPRCM compatibility and institutional preferences.

The third case shows a storyline approach. An extreme event, the hot Scandinavian summer in 2018, is analysed regarding its relation to the changing climate but also its place in different climate model ensembles uncertainty space. The case study shows the importance of larger ensembles and to not neglect extreme members of an ensemble and that sensible and informed selection processes are necessary.

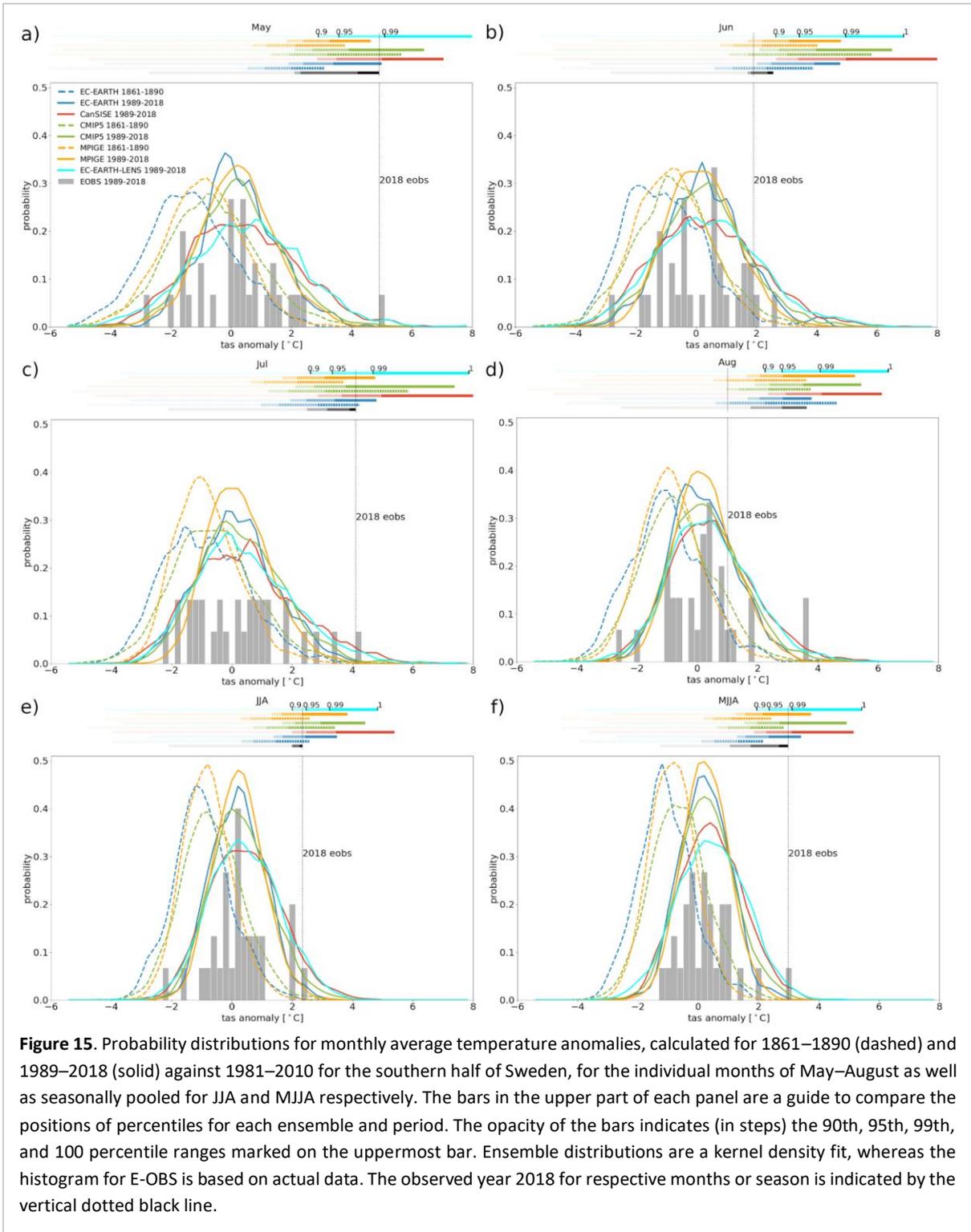


Figure 15. Probability distributions for monthly average temperature anomalies, calculated for 1861–1890 (dashed) and 1989–2018 (solid) against 1981–2010 for the southern half of Sweden, for the individual months of May–August as well as seasonally pooled for JJA and MJJA respectively. The bars in the upper part of each panel are a guide to compare the positions of percentiles for each ensemble and period. The opacity of the bars indicates (in steps) the 90th, 95th, 99th, and 100 percentile ranges marked on the uppermost bar. Ensemble distributions are a kernel density fit, whereas the histogram for E-OBS is based on actual data. The observed year 2018 for respective months or season is indicated by the vertical dotted black line.

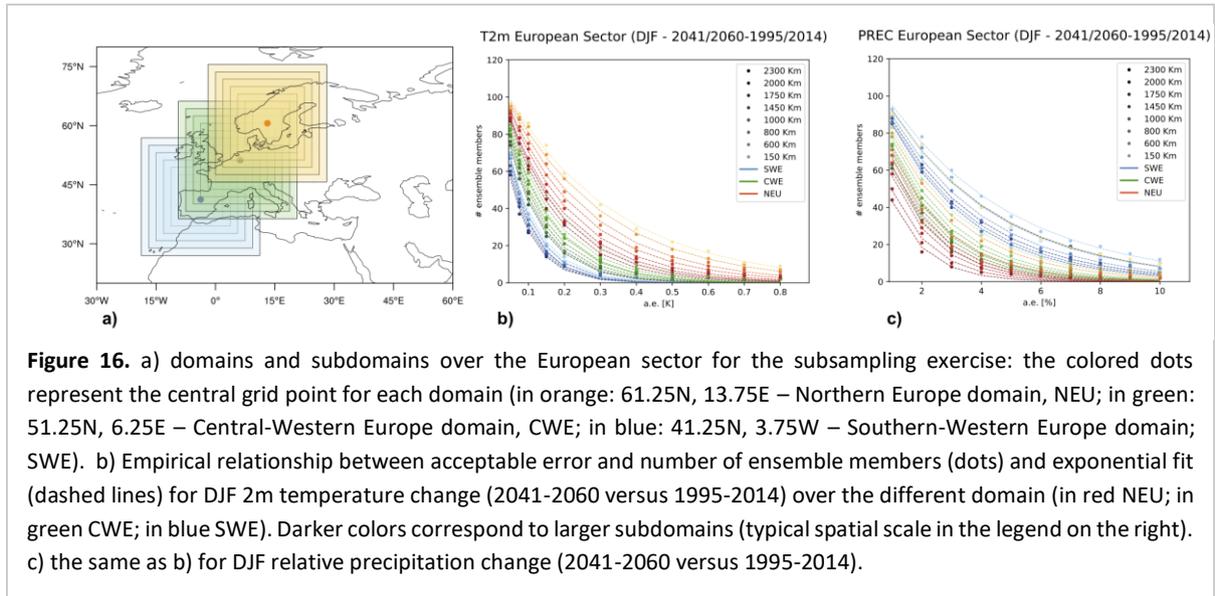
3.6 Science part 5 - Quantifying the minimum number of ensemble members (CMCC)

INTRODUCTION -- How large an ensemble of climate simulations needs to be for different purposes is a challenging issue within the modeling community. We take advantage of a single model initial-condition large ensemble (SMILE), to investigate the number of ensemble members needed to characterize a given climate signal or to fully represent internal variability. The recent literature shows clearly how the needed ensemble size generally depends on the application of interest. In a climate change perspective, Deser et al. (2012) report the minimum ensemble size required to reproduce a statistically significant change compared to present-day climate in the global average temperature and precipitation: in the first decade less than 10 members are needed for temperature, while for precipitation 40 members are not enough. Maher et al. (2018) estimate that at least 30 ensemble members are needed to robustly estimate El Niño Southern Oscillation (ENSO) variability. Looking at the stratospheric response to volcanic eruption, Bittner et al. (2016) report that the number of ensemble members needed to capture changes in the 10 hPa zonal wind after a strong eruption crucially depends on latitudes, with much more members needed at high latitude (around 40) than at the southward flank of the jet (less than 10). These are just a few examples, clearly showing how the number of members required to represent a given process fundamentally depends on the intensity of the signal and/or on the related internal variability, which are properties intrinsic to the considered phenomenon. Here we will present two different approaches for the estimate of the proper ensemble size to represent a given climate signal within a certain range of uncertainty, and we will provide some case studies. The general idea is to subsample the full ensemble in order to characterize its properties as a function of its size. A specific focus will be placed on how this estimate varies with a varying extent of the considered spatial domain. The aim behind this study is twofold: whilst it may enhance the exploitation of the existing climate prediction products, helping end-users to define the optimal (sub)ensemble size needed for a specific purpose (e.g., to study climate variability on a specific target area, or to define the proper input for a tailored impact study), it also provides an informative framework for climate scientists (e.g., in the design phase of new experimental efforts).

METHOD 1 (SUBSAMPLING APPROACH) -- The MPI-Grand Ensemble (MPI-GE; Maher et al., 2019) is a 100-member ensemble from the MPI-Earth System Model in the low-resolution setup (MPI-ESM1.1), forced with the CMIP5 simulation protocol in the historical period (1850-2005) and in the scenario period (2006-2099). In this framework, for the scenario period, we adopt the RCP8.5 concentration pathway. By construction in a SMILE all the members include the same external forcing and have been produced by the same numerical model (i.e. no structural or scenario uncertainty is considered), hence the differences across them are due to the intrinsic chaotic nature of the climate system. The full ensemble provides the best estimate of the probability distribution characterizing a given variable or field: our goal is to study how many ensemble members are needed to retain the same information. In other words, we want to test the sensitivity to the ensemble size of the estimate of the distribution parameters. To do so, an iterative subsampling is applied: for each subsample size ($k = 2, 3, 4...99$) an extraction without replacement is performed N times ($N = 1000$), obtaining N couples, N triplets and so on. As emphasized by Milinski et al. (2020), who recently proposed an analogous approach, the choice to resample without replacement implicitly reproduces what happens when a limited number of ensemble members is produced running a set of model simulations. For each subsample size, the subsampling procedure generates N synthetic ensembles over which the different ensemble distribution parameters may be computed, obtaining, for example, N estimates for the ensemble

average, for the ensemble variance and so on, which will be proper of the considered subsample size. From these N estimates and for each subsample size k , a distribution of values can be defined, the spread of which is a measure of the error in taking the subsample of size k instead of the full ensemble. To test our confidence in adopting a reduced ensemble size to estimate the chosen parameter, we can compare the spread of these distributions (i.e. the 5th - 95th percentile range), with some arbitrary threshold, representing the acceptable error for the user, given the specific instance under examination. In this framework, this analysis is applied looking at climate signals averaged over a specific domain. Different regions and scales are typically affected by different degrees of variability, impacting in turn the requirements in terms of ensemble members even when analyzing the same signal (e.g. the forced response due to anthropogenic forcing). In order to obtain some indication of how the spatial aggregation influences the ensemble size requirement, starting from the chosen target region a set of subdomains is defined over which the field of interest is averaged, and the subsampling analysis is repeated.

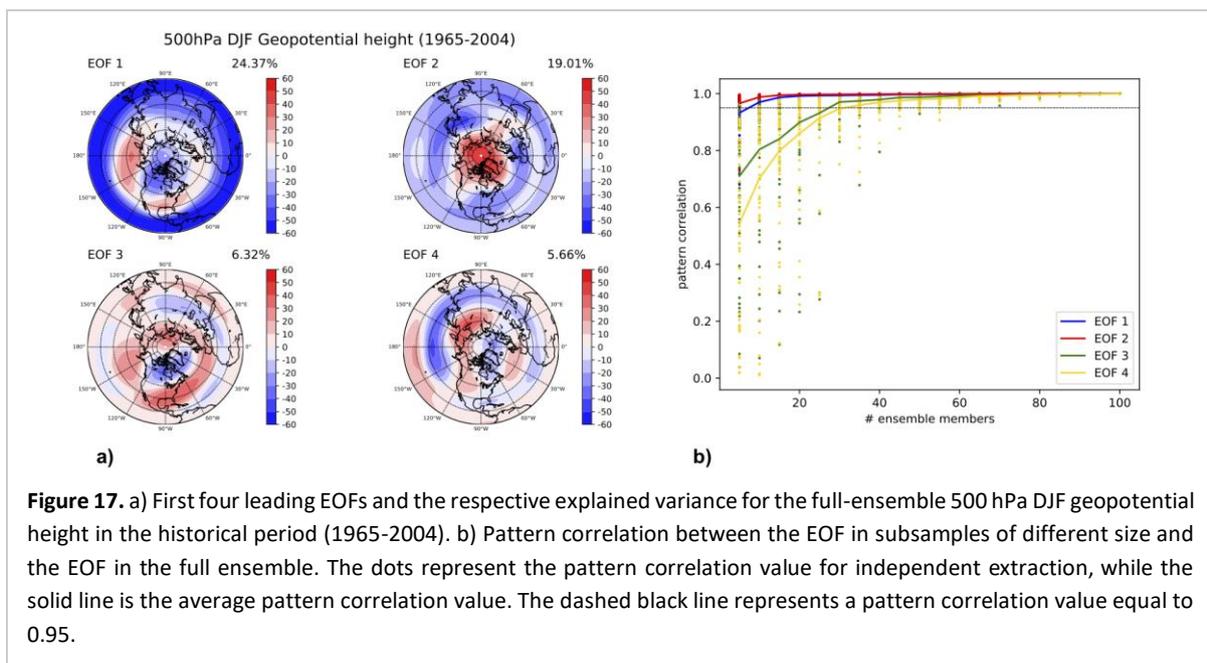
ENSEMBLE SIZE AND SENSITIVITY TO THE SPATIAL SCALE -- a case study -- The proposed approach has been adopted to evaluate the number of ensemble members needed to characterize the climate change signal over the Euro-Mediterranean domain. In particular, the focus is on the DJF 2 m temperature and precipitation change over the period 2041-2060 compared to present day climatology (i.e. 1995-2014). The target domain is reported in Fig. 16a: three different sectors have been evaluated, centered over selected grid points (61.25N, 13.75E; 51.25N, 6.25E; 41.25N, 3.75W) in Northern, Central-Western, and Southern-Western Europe. For each target domain, the subsampling exercise has been performed, allowing to define how many ensemble members are needed to reach a spread of the subsample distribution comparable to a given acceptable error. By means of a set of decreasing values for the acceptable error, it is possible to build an empirical relationship linking the desired level of uncertainty with the needed ensemble size (Figs. 16b,c). The curves show an exponential-like behavior as emphasized by the exponential fit plotted over the empirical points for both absolute changes in temperature (Fig. 16b) and relative changes in precipitation (Fig. 16c). Here different colors identify different sectors, while the shading is used to label different spatial scales. Generally, the larger the domain, the smaller the number of ensemble members required to reach the same acceptable error. The regional behavior crucially depends on the variable of interest: while the ensemble size needed to reach the same acceptable error decreases moving southward for temperature, the opposite holds for precipitation. For 2m temperature, the bundles of curves for the different regions are clearly separated in clusters of different shape, proxy of the different variability affecting the regional climate change signal (i.e., the less ensemble members are needed to reach a certain acceptable error, the higher the agreement across the ensemble – as shown for example for the southern sector compared to the northern one). On the contrary for precipitation a less evident clustering can be noticed. Indeed, at finer spatial scales the precipitation signal is generally characterized by a reduced spatial coherency and a more pronounced variability regardless of the region considered.



METHOD 2 (ENSEMBLE MODE DECOMPOSITION) -- In the following we will present a methodology to detect how many ensemble members are needed to represent the same modes leading the variability in the full ensemble, always taking advantage of the MPI-GE. The Empirical Orthogonal Function (EOF) analysis is a widely used approach to detect and separate spatial modes of variability and their temporal variation in a dataset (e.g., Hannachi et al., 2007). The input data are typically organized in a data matrix where each row corresponds to a time series (i.e., the temporal behavior of the field) for each grid point. From such data matrix, once properly scaled and standardized, it is possible to derive the covariance matrix for the field; its eigenvectors and eigenvalues provide respectively the orthogonal mode of spatial variability and a measure of the amount of variance explained by each mode. It is important to notice that by construction EOFs are orthogonal, implicitly hindering the direct interpretation of the obtained patterns as real physical modes. Nevertheless, the canonical EOF analysis is commonly considered as a powerful method for the analysis of huge datasets and for dimensionality reduction. In this framework we want to extend the canonical EOF approach and apply it over the entire ensemble. The ensemble data matrix is then defined by aligning in each row (i.e. for each grid point of the considered domain) the time series for each ensemble member. If we denote with N the size of the full ensemble, with S the number of grid points, and with T the number of available time steps, the data matrix defined in this way has dimension $S \times NT$. Anomalies have been previously computed subtracting to the full field its own climatology for each ensemble member, and the data matrix is normalized dividing these anomalies at each grid point by the standard deviation computed over time and over the ensemble. The EOF modes computed from such a data matrix describe the variability included in the full ensemble. Once this reference result is achieved, we can adopt an iterative subsampling to check how many ensemble members are needed to obtain the same set of modes. In other words, the ensemble is subsampled extracting randomly without replacement an increasing number of members from the full ensemble, and for each subsample the EOF analysis is repeated. In this way a set of spatial patterns for each subsample size is obtained, comparable to the one computed for the full ensemble.

ENSEMBLE SIZE AND MODE OF VARIABILITY - a case study -- Here we show an example, looking at the modes of variability of the winter (DJF) 500 hPa geopotential height in the historical period (1964-2005). Our benchmark is represented by the full ensemble variability, and here we will focus on the

first four modes (Fig. 17a), explaining around 55% of the total variance. The first EOF shows a wavy signal crossing the North Pacific and the North American continent, resembling the Pacific North American pattern (PNA e.g., Wallace and Gutzler, 1981). The second mode presents a clear annular structure with perturbations of opposite sign over the polar region and in the surrounding zonal ring around 45N. This pattern can be interpreted as the tropospheric counterpart of the Arctic Oscillation (Thompson and Wallace, 1998), generally defined as the leading EOF of the wintertime sea-level pressure field. The higher order modes are characterized by wave-train patterns crossing the whole hemisphere, more difficult to be interpreted as potentially affected by the orthogonality constraint implicit in this kind of decomposition. Once the full ensemble EOFs have been defined, the subsampling has been performed, and the EOF modes have been computed for a set of random subsamples of increasing size. The aim is to find the minimum number of ensemble members needed to represent the same modes of variability included in the full ensemble. The results of this subsampling exercise are summarized in Fig. 17b. To compare the full ensemble modes with the analogous ones derived from subsamples of different size the pattern correlation has been computed. Since the EOF patterns are defined up to sign, the absolute value of the pattern correlation is plotted. To reduce the potential impact of the subsampling choice, the random extraction is repeated 50 times for each sub-ensemble size and for each of these repetitions the pattern correlation value is computed (reported as dots in Fig. 17b). The average behavior of the pattern correlation (solid lines in Fig. 17b) shows that the first two modes converge very rapidly to the full ensemble value, with a pattern correlation greater than 0.95 with 5 members. In contrast, around 30 ensemble members are required to represent the higher order EOFs (i.e., to have pattern correlation over the 0.95 threshold for both the third and the fourth mode). It should be noticed (not shown) that the amount of explained variance for the four EOFs remains constant across the different subsample sizes. This result, together with the behavior of the pattern correlation, reveals that the subsampling does not revert the separation of the modes.



FINAL REMARKS -- Two different methods have been presented, providing an assessment of the minimum number of ensemble members needed to characterize the average signal over the region of interest or to properly represent complex structures in space and time (i.e., the EOF modes). In the first case, the strength of the proposed subsampling approach is that it can be easily generalized and

adapted, taking into account how the subsampling impacts on different distribution parameters (e.g., higher order momenta, standard deviation etc.). In the example reported here the climate change signal, defined as changes in the ensemble mean, has been examined obtaining as result a pronounced regional and variable dependence. Likewise, the same approach can be applied to some measure of ensemble spread (e.g., the ensemble variance), which by construction can be considered as a proxy for internal variability, when SMILEs are taken into account. Recently a similar methodology has been presented by Milinski et al. (2020); in this framework we have extended the analysis by considering how the extension of the spatial domain over which the signal is defined affects the choice of the ensemble size. In this sense, the choice of the target domain and of the relative subdomains is crucial and depends on the physical properties of the signal of interest. For example, when looking at the tropical Pacific SSTs or at the North Pacific jet (not shown), we have adopted a zonally elongated target region and subdomains of gradually reduced zonal extension, but constant meridional one. In the second instance, the definition of the EOF modes over the ensemble allows to focus on more complex signals and hence to retain more physical information, with the due caution in the interpretation related to the intrinsic orthogonality constraint. It can be demonstrated that the covariance matrix obtained from our choice in the definition of the data matrix is equivalent to the ensemble covariance matrix in Wills et al. (2017), where the ensemble EOFs are defined as the starting point for the computation of patterns of low-frequency climate variability and change. The example presented here focuses on the Northern Hemisphere large-scale winter circulation patterns: while the first two modes are extremely stable across the subsampling, the higher order EOFs show an enhanced variability, as can be noticed from the spread of the dots in Fig. 17b. However, the average pattern correlation converges relatively fast. Finally, it should be noticed that there is the potential to generalize this approach to overcome the well-known limitations implicit in the EOF analysis, by extending to the ensemble more general mode decomposition techniques (e.g., the rotated EOF).

3.7 Science part 6 - Pseudo-global warming experiments to create event sets of future weather (KNMI)

INTRODUCTION--The impact of recent west-central European drought events on society and nature once again triggered questions regarding the role of climate change in the occurrence and extremity of such events and on what to expect under continuing global warming. While the probability of heat waves in this region is demonstrated to have increased in response to anthropogenic climate change, the attribution of extreme drought conditions is profoundly difficult by the complexity of processes contributing to wide-spread drought conditions.

In the contribution presented here the impact of global warming on soil moisture drought severity in west-central Europe is investigated by employing so-called pseudo-global warming experiments, which set the 1980-2020 period against the background of a globally warmer world. The “future analogues” of present-day drought episodes allow us to investigate changes in drought severity conditional on the day-to-day evolution of the atmospheric circulation in the historical period. The purpose is to provide robust, physically consistent scenarios of what global warming entails for extreme droughts like 2018, and for the full range of wet to moderately dry years in the historical period. As such this approach is intended to complement projections of changes in drought risk derived

with the conventional climate modelling approaches, giving an explicit reference to collectively experienced real world events.

METHODS--The simulations are performed with the RCM KNMI-RACMO2 at 12 km resolution. External forcings for aerosols and greenhouse gases have been implemented according to CMIP5 prescriptions. The model domain is centered over west-central Europe, and covers most of Europe. The analyses are based on two sets of RCM simulations: a present-day simulation (REF), driven by ERA5 re-analysis fields, and pseudo-global-warming (PGW) simulations corresponding to a 2 K global warming. Both sets consist of a climate run covering the period 1980-2020 (climREF and clim+2K). The PGW-simulation is essentially a rerun of the REF-simulation, but with perturbed lateral boundary conditions (surface pressure and 3-dimensional temperature, humidity, and wind) and perturbed sea surface temperature. The perturbations are determined from global climate model (GCM) projections as the monthly mean climate change signal in the 30-year period in which the target global warming level is reached with respect to present-day conditions (1991-2020). With that a large part of the climate signal is captured, but the day-to-day evolution of the synoptic-scale circulation in the PGW simulations remains essentially determined by the reanalysis forcing and is therefore very similar to the sequence seen in the present-day (or control) simulation.

Since climate models differ in their regional climate response, we have applied perturbations from three different GCM initial-condition ensembles: a 16-member EC-EARTH v2.3 ensemble produced at KNMI, a 4-member HadGEM2-ES ensemble and a 3-member MPI-ESM-LR ensemble from the CMIP5 archive, referred to as, respectively, EC, Had and MPI. The perturbations are derived from initial-condition GCM ensembles rather than from a single simulation per GCM to obtain a robust estimate of the forced climate response. All GCM ensembles are run under the high-end emission scenario (RCP8.5). Fig. 18 shows some features of the perturbations for the three GCMs. Obviously there are large similarities in climate change response between the three ensembles, but details like e.g. the response in pressure gradient and the shape of vertical temperature response are different.

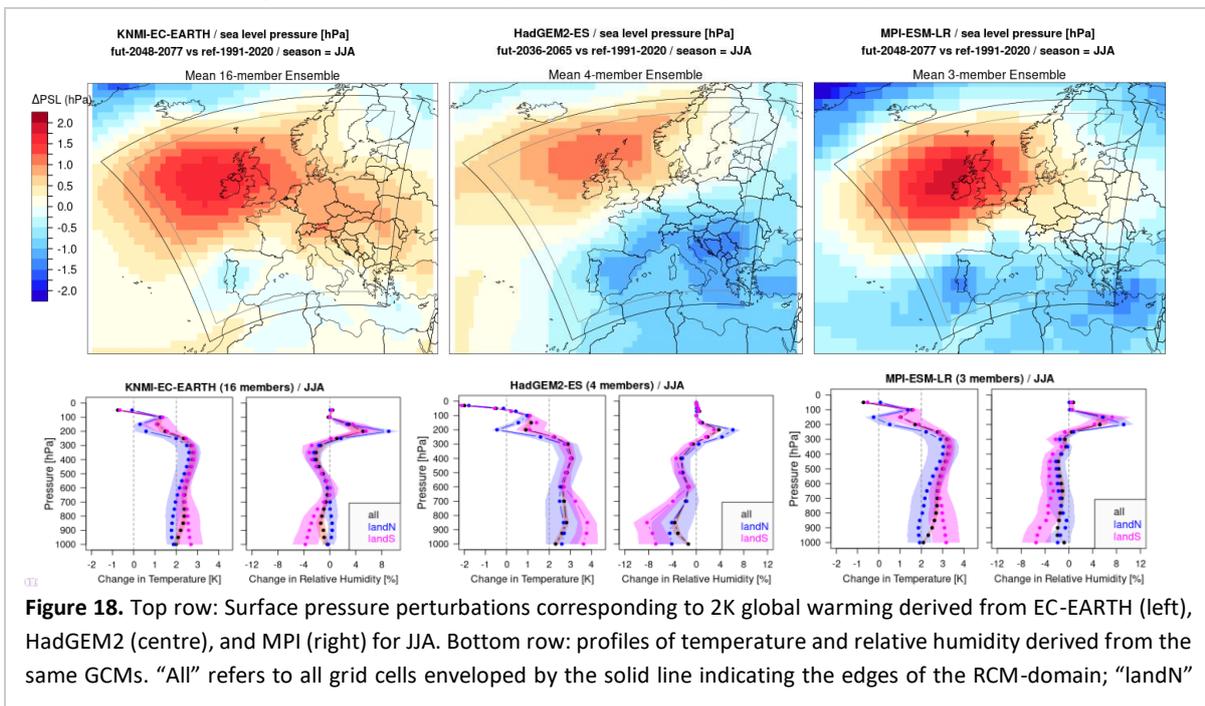


Figure 18. Top row: Surface pressure perturbations corresponding to 2K global warming derived from EC-EARTH (left), HadGEM2 (centre), and MPI (right) for JJA. Bottom row: profiles of temperature and relative humidity derived from the same GCMs. “All” refers to all grid cells enveloped by the solid line indicating the edges of the RCM-domain; “landN”

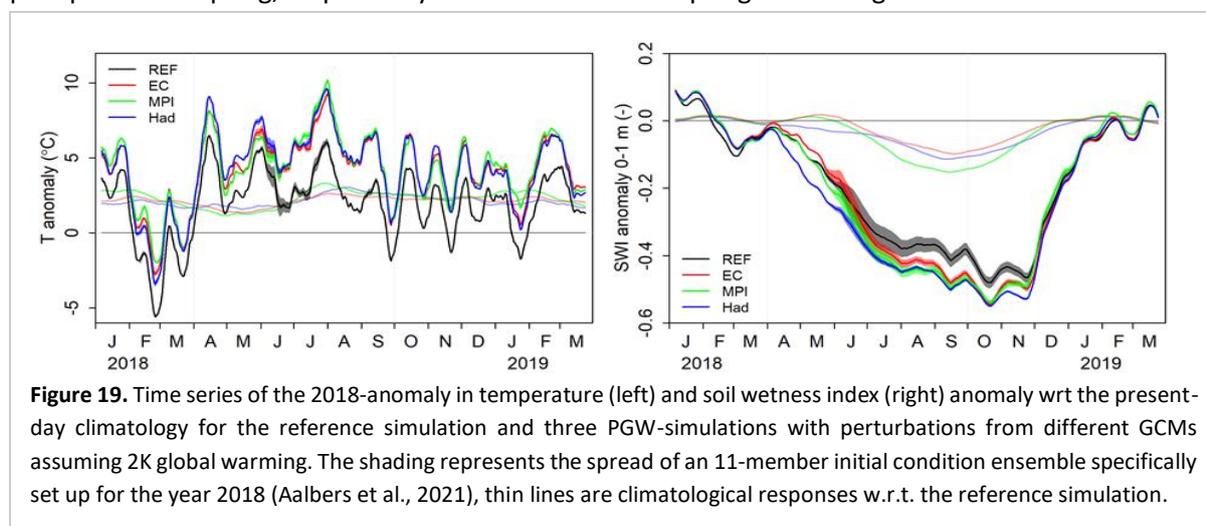
refer to land points north of 50°N; “landS” to cells in between 35 and 50°N. The shading indicates the spread across the regions.

Hereafter, soil moisture drought conditions in the model are based on the exceedance of a seasonally varying threshold of the soil wetness index (SWI) of the top 1 m of the soil, where SWI is defined as the soil moisture availability scaled between field capacity and permanent wilting point. A soil moisture drought event is defined as the consecutive period in which SWI is below the 5th percentile determined from the reference simulation. The drought severity is expressed in terms of the drought deficit volume (units: mm d), which integrates drought duration and drought intensity. It is calculated as the cumulative SWI deviation from the 5th percentile threshold over the drought episode. The drought intensity is defined as the drought deficit volume divided by the drought duration.

The analysis is carried out for a designated region consisting of the larger river basins in west-central Europe that discharge in the North Sea, namely the Rhine, Meuse, Scheldt, Ems, Weser and Elbe. These river basins are part of the area where the 2018 soil moisture drought episode (Fig. 19) was most severe and lasted longest.

MAIN RESULTS—A comprehensive analysis of the 2018 west-central European drought event and its projection in a warmer climate within the framework of PGW is discussed in Aalbers et al. (2021), included in Appendix A4. Here we present results for the 2018 drought event in the context of the reference period 1980-2020 both for present-day conditions and for future conditions in a +2K globally warmer world under PGW.

As seen in Fig. 19, the 2018 drought onset and ending under global warming occurs at roughly the same time as for the present-day event. The drought duration under PGW is thus hardly affected or even decreases, while the drought intensity shows a moderate increase (8 mm or 22%), resulting in an increase in the drought deficit volume of 20% under 2K warming. The Had and MPI-perturbed simulations show stronger increases in drought severity (+39%, resp. +25%) owing to decreasing precipitation in spring, respectively weaker increases in spring and stronger decreases in summer.



Yet, there are summers in the 1980-2020 period with much stronger responses in soil moisture than 2018. This is illustrated in Fig. 20 showing the drought intensity and duration for all years in the 1980-2020 period for the reference simulation and each of the PGW-simulations. Under present-day

projections. The advantage is that results are less sensitive to GCM biases in the reference climate state and that, to a large extent, the effect of mean climate change is captured. As such, changes in droughts can be directly related to real world events and their societal impact, which make the results very tangible and therewith useful for climate change communication.

Under 2K warming almost all years in the 1980-2020 period show a decrease in soil moisture availability in (spring), summer and autumn, consistent with results based on transient climate model simulations (next section). Compared to the change in the climatology, the soil moisture response in west-central Europe under the circulation of 2018 is relatively modest, primarily because evaporation in 2018 is already strongly moisture-constrained in the present-day simulation. Also, the increase in (potential) evaporation owing to much higher temperatures is partially compensated by an increase in precipitation. The 2018 soil moisture deficit volume in the PGW-experiments increases, primarily due to the increase in drought intensity. The drought duration is barely impacted owing to increasing precipitation in spring, autumn and winter. It is emphasized that the increase in drought severity in 2018 co-occurs with an increase in local summer temperature which is considerably larger than 2K global warming.

Finally, a very interesting outcome is that the response in temperature and soil moisture drying can be particularly large for years with moderately dry conditions in the present-day climate, implying that years that went unnoticed under present-day conditions may emerge as very dry years in a warmer climate. Using present-day thresholds, the drought frequency strongly increases under +2K warming, with 2019/2020-like deficit volumes occurring on average every three years, 2003-like deficit volumes every eight years, and under strongly enhanced temperatures. This shows that even without taking into account changes in the frequency of atmospheric blocking conditions, the probability for severe drought events in west-central Europe is strongly enhanced.

3.8 Science part 7 - Selecting weather events from initial condition large ensembles (KNMI)

INTRODUCTION-- In the immediate aftermath of an extreme weather or climate event, from different parts of society questions arise on the role of climate change in creating this particular event and what future events might look like taking into account further climate change (Van Oldenborgh et al. 2021). These questions matter from a societal perspective, as the answers inform people on the consequences of climatic changes on local weather, but can also inform policy makers on whether adaptation measures are required to cope with potentially more severe weather events that may occur because of climate change.

Event-based scientific research, motivated by a specific observed or theoretical weather event, can provide answers to these questions. Such research can take many forms, but all benefit from the fact that the human mind is more impressionable to concrete and relatable information than general information (Shepherd et al. 2018). For climate information this means that information on specific future events, especially when related to observed events in the recent past, is better received and more actionable than general climate scenarios. 'Storylines' of events that could be in a warmer future are thus of societal use; hence the development of scientific methods for creating such storylines is of interest.

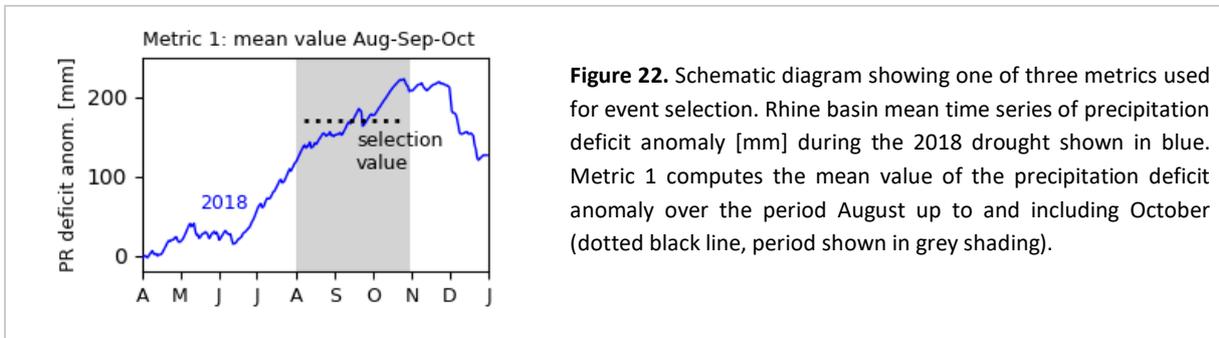
To transpose an observed extreme event to a warmer future, a few climate modelling routes are possible. The most common is to perform a nudging or pseudo-global warming (PGW) experiment, in which the model is constrained to reproduce for atmospheric circulation of the event (see e.g. science part 6 of this deliverable), however, it is unclear what is the best way of perturbing thermodynamics or boundary conditions (Silman et al. 2019). Alternatively, comparable events can be selected from unconstrained model experiments. Here we describe an explorative study on a new method for creating physical storylines of possible future events that are similar to an observed weather event of choice using this second ‘selection’ approach. What follows is a conceptual description of the proposed methodology, a proof-of-concept where we apply this method to the extreme north-western European drought of 2018, and some considerations of this method in relation to existing ones. The peer-reviewed journal article describing this study is included in Appendix A5.

METHODS-- From a large collection of similar runs from a single climate model (i.e. a Single-Model Initial Condition Ensemble’, SMILE), different runs sampling different phases of climate internal variability, we aim to select events that resemble the observed event in a pre-defined way. If we can satisfactorily do this for the present-day climate, and we find enough comparable simulated events to the observed event, we can assume the same selection metrics or methods would work in future climates as well. The selected simulated events can then be compared across climate states, providing information of the influence of climate change on this (type of) event(s).

To ensure robust and reliable information is generated, there are a few things to test and consider. Does the climate model realistically represent the local/regional climate, including the event of interest? An assessment of model biases is required, though often difficult for the particular extreme event of interest given the rarity of occurrence and hence limitations of observed data. Second, the ensemble of data from which events are selected should be sufficiently large, such that the selected simulated events are comparable in terms of return period and severity to the observed extreme event. This is important, as the climate change response can be different for events of different return periods (Van der Wiel et al., 2018, 2019). Lastly, we note that each event is different in its specifics, no two heat waves are exactly the same. This negatively impacts the calculation of the climate change response, which can be clouded by event differences due to natural variability. To circumvent this, at least partially, we take a composite mean over a few selected simulated events. The composite mean event, referred to as an ‘event analogue’ here, is less influenced by natural variability which improves the robustness of the calculation of the impact of climate change on the event (see also section 3.6 on internal variability).

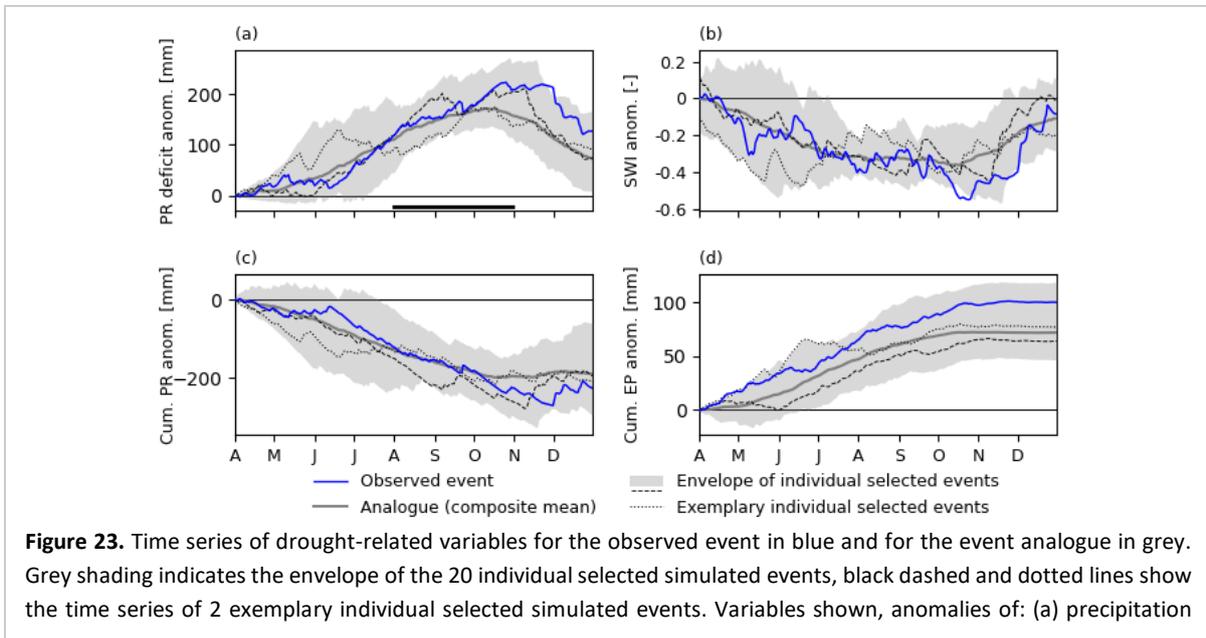
PROOF OF CONCEPT-- We apply the above method to study what droughts like the north-western European drought of 2018 would look like in warmer future climates. The summer of 2018 was exceptionally dry and warm in north-western Europe (Philip et al. 2020; Zscheischler and Fischer, 2020), this negatively impacted nature and many societal sectors (e.g. agriculture, shipping, drinking water, housing). Here we aim to investigate what a drought event like 2018 would look like in a warmer future climate. We create ‘storylines’ of future events that are comparable to the observed event, which contribute to answering user questions like “How much worse can droughts get (e.g. more severe, more widespread, longer in duration)?” and will help policy-makers in the planning of adaptation measures. Focus will be on meteorological drought processes.

We use a set of three large ensemble experiments from the EC-Earth (v2.3; Hazeleger et al. 2012), each consisting of 2000 years of simulated weather for a given state of global climate (present-day, pre-industrial +2 °C warming, and pre-industrial +3 °C warming; Van der Wiel et al. 2019). Three metrics (e.g. Fig. 22), each describing a different aspect of the observed drought event are defined, and used to select the 20 most similar simulated events in each ensemble (3x 20 out of 2000 summers). The simulated analogues of the observed drought of 2018 are then computed by taking the composite mean of the 20 selected events.



In Fig. 23 we compare the constructed analogue from simulated climate model data to the observed event. For the variables shown here (and others, see Appendix A5) the observed event lies within the range of selected events. Consequently, the analogue reproduces the event quite well. Lower than normal precipitation (Fig. 23c) and higher than normal evapotranspiration (Fig. 23d), resulting in a high precipitation deficit (Fig. 23a, meteorological drought) and low plant available soil moisture (Fig. 23b, agricultural drought). This was all driven by the presence of persistent high pressure systems, high temperatures and sunny conditions.

Given the high degree of similarity between the observed event and the analogue (Fig. 23), we have confidence that we can use the analogue approach in conjunction with the EC-Earth large ensembles for a study of large-scale droughts in western Europe. We will therefore continue to investigate the effects of climate change on potential future events.

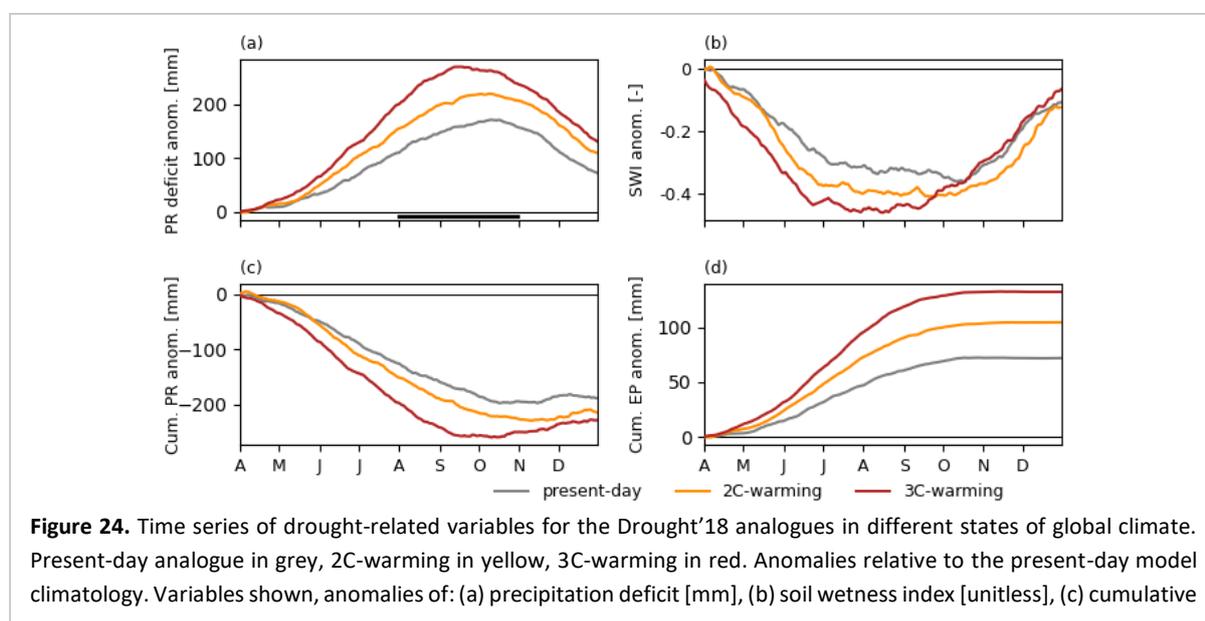


deficit [mm], (b) soil wetness index [unitless], (c) cumulative precipitation [mm], (d) cumulative potential evapotranspiration [mm]. The horizontal black bar in (a) shows the event selection period, as shown in Fig. 22.

The future event analogues (Fig. 24) show that comparable events have higher precipitation deficits in 2C- and 3C-warmer climates (Fig. 24). This drought intensification is caused by a decrease of precipitation and an increase of evapotranspiration (Fig. 24c,d), and leads to a further drying out of soils (Fig. 24b). The intensification of droughts in future climates is found independent of the method of event selection, i.e. all metrics show this intensification (Appendix A5). However, the precise timing and magnitude of the changes does depend on the method of event selection. For the metric shown here (based on high Aug-Oct precipitation deficit), we find that a dry spring kick-starts the drought and is temporally compounded by a dry summer. Taken together, this leads to the highest late summer deficits in the ensemble dataset. Another metric, focussing on midsummer dry conditions, does not show a signal in spring, in that case drought intensification is caused by exceptionally low amounts of rain, warm temperatures and sunny skies in June, July, August (Appendix A5).

The changes in drought severity are larger in a world with 3 °C warming than in a world with 2 °C warming. Climate change mitigation will thus help to limit the societal and natural impact of future drought events somewhat. Societal adaptation to more severe drought events is inevitable if impacts are to be limited to present-day levels.

COMPARE TO PGW METHOD-- In the previous science part of this report (section 3.7) a PGW experiment for the same observed drought event was described. Both methods transfer an observed extreme event into a warmer future, creating tangible climate change information for users and policy makers. And despite the fundamental difference in the two approaches (PGW recreating the exact circulation patterns as it occurred, with monthly mean changes in atmospheric circulation, but assuming circulation changes on top of those have a minor role; whereas selecting from a SMILE aims to find similar impacts, not necessarily forced by the exact observed atmospheric circulation, but allowing all forced changes in atmospheric circulation), the results are qualitatively comparable (e.g. Figs. 19 and 24b).



precipitation [mm], (d) cumulative potential evapotranspiration [mm]. The horizontal black bar in (a) shows the event selection period. Note all lines are composite means over 20 selected simulated events, as grey lines in Fig. 23.

SUMMARY -- The aim of this work was to develop a method for the creation of event-based storylines of future extreme events similar to an observed weather event, by means of selection from large ensemble climate model simulations. By means of a case study, we show that this method has promise, if the ensemble of choice is of sufficient size to include a couple of comparably extreme events to the observed event. More details on the method and results of the case study can be found in Appendix A5, which includes a discussion of the balance between increasing event return period and reducing noise.

It should be investigated if the proposed approach can be extended to other types of extreme events, for example heat waves and large scale extreme rainfall events. Furthermore, for users the translation to weather impacts (for drought, e.g. reduced crop yields, drinking water quality and quantity issues, irreversible harm to nature reserves) is important, the value of physical climate storylines for this community has not yet been fully explored.

3.9 List of related EUCP publications

- Peer-reviewed publications, other reports [all finished]
 - Merrifield, Brunner, Lorenz, Medhaug, Knutti, 2020. An investigation of weighting schemes suitable for incorporating large ensembles into multi-model ensembles, *Earth System Dynamics*, 11, pp.807-834.
 - Palmer, Booth, McSweeney, 2021. How does the CMIP6 ensemble change the picture for European climate projections?, *Environmental Research Letters*, 16(9), p. 94042.
 - Van der Wiel, Lenderink, De Vries, 2021. Physical storylines of future European drought events like 2018 based on ensemble climate modelling, *Weather and Climate Extremes*, 33, pp.100350.
 - Wilcke, Kjellström, Lin, Matei, Moberg, Tyrlis, 2020. The extremely warm summer of 2018 in Sweden – set in a historical context, *Earth Syst. Dynam.*, 11, 1107–1121..
- Planned future publications [submitted, in preparation, in plans, etc]
 - Aalbers and co-authors. The 2018 west-central European drought projected in a warmer climate. *In preparation*.
 - McSweeney and co-authors. Evaluating large-scale circulation for Europe in a coupled PPE. *In preparation*.
 - Sexton, McSweeney and co-authors: Describing future UK winter precipitation in terms of changes in local circulation patterns. *In preparation*.
 - Benassi and co-authors. Sensitivity of climate signals to the ensemble size. *In preparation*
 - Wilcke, Belusic and co-authors. Setting convection permitting regional simulations into the broader RCM and GCM context. *In plans*.
- Storyboard science summaries for wider communication:
 - We are in the process of developing online storyboards for some of the methods described here. <https://eucp-project.github.io/usecases/>
- Open data or scripts:
 - ClimWIP in ESMValTool
https://docs.esmvaltool.org/en/latest/recipes/recipe_climwip.html
 - Drought'18 storylines data <https://zenodo.org/record/5083160>

3.10 Discussion

This deliverable reports on various scientific projects that either develop new methods for subselection or build on existing methods to improve and investigate their fidelity. Here we note some common findings and recommendations.

Both Section 3.2 (ClimWIP method) and Section 3.3 (qualitative framework) note that scientific thinking in the climate science community has shifted from a historic model-democracy approach (one model one vote) to an approach taking into account the individual performance of a model and potential dependencies between models. It is further noted that specific model performance is highly-dependent on the exact use case and metric of choice. This shows the importance of co-production projects, in which the process of model selection for a particular process and the resulting sub-set of models are specifically designed for the process of interest. Also Section 3.5 notes “*This dialog and*

building of understanding is essential for a user tailored product.” The three sections provide different approaches to subselection of models, though we note they share some of the same principles, e.g. hierarchical clustering procedures or the aim to keep the variability of the original ensemble in the resulting subset.

The second aim of this deliverable is the development of event sets of present and future extreme events. Such sets can work towards creating better understanding of future climate change projections, because the outcomes are more tangible to users and relate directly to experienced weather events. Sections 3.7 and 3.8 describe two independent approaches, one by doing targeted RCM experiments for the event of interest, the other by sub-selection from large ensemble simulations. The targeted experiments, using the PGW approach, are especially valuable if there is a limited number of events of interest, and if circulation changes for the event specifically (on top of seasonal mean changes) are expected to be small. If computational constraints make such experiments impossible, or if the potential study is more explorative, the sub-selection approach can provide a less-computationally intensive alternative. Internal variability in large ensemble simulations (see also section 3.6) is then used to select the analogue simulated events. Also here, as with the subselection of models, co-production of event sets with users is an obvious next step to improve potential usefulness.

Finally, many interlinkages can be found between the presented methods. We will note one possible connection. The axes of climate projection uncertainty identified using the methods of Section 3.4, can be used to create storylines or scenarios of future climate change. These scenarios can then be applied to the PGW method of Section 3.8 to create sets of future weather events that are physically consistent with the scenarios of the previous set. In this, and other ways, the subselection methods can inform further activities which use ensembles of models or ensembles of experiments, to add value to data and ultimately guide the process to create user-oriented actionable climate information.

4. Lessons Learnt and links Built

Synergies and links with other deliverables, work packages, and projects:

- Synergies exist between this task and WP5, most notably WP5’s work on storylines. We have been in regular contact, discussing for example how to add-value to our ‘academic’ storyline exercise, by including users in a co-production effort. Also, we connected with the EUCP multi-user forum, the users (e.g. Anglian Water) were interested in our storyline approaches to understand possible future drought events. Collaborative continued work to make impact calculations has started.
- In WP3, the spatial merging methods draw on the principles of sub-selection discussed in this report.
- Outside EUCP, members of WP2 have worked with and advised the EURO-CORDEX community on their ongoing work on sub-selecting from the CMIP6 model ensemble for new Euro-CORDEX simulations. EUCP WP2 has informed this activity directly through both the work on the assessment on models for sub-selection (Section 3.3 of this report) and model families (Section 3.2). WP2 work on constrained ranges has also been drawn on to inform the sub-selection of models with respect to the plausibility of observed trends (Ribes et al. 2021). This

activity represents a significant step forward in the CORDEX community and will lead to more useable EURO-CORDEX projections which better represent the spread in driving global models.

Lessons learnt in the process of this task:

- By finalizing outputs well before deadlines, one leaves enough room to identify synergies between tasks and work packages. Though we have been very productive within our institutes, we have missed face-to-face meetings, which has meant that collaborative work across the consortium was made more difficult.
- Similarly, parallel activities in work packages make it harder to have information from one piece of work to feed into another bit of work.
- The pieces of work described in this deliverable have an academic focus, it would have been great had we been able to add a user focus. Such co-production work adds further value, but was hampered by the inability to organize in-person workshops.
- Finally, the pandemic severely limited in-person meetings, making it more difficult to form connections and collaborations across institutes. Bi-weekly online coffee meetings, with both time for informal chat as for scientific discussion, helped create a team of people within the work package. We recommend having such regular meetings, even post-pandemic when normal meetings are possible again, it helped us stay in contact and kept communication lines short.

5. References

- Aalbers and co-authors, 2021. The 2018 west-central European drought projected in a warmer climate, in preparation.
- Allan R, Ansell T (2006) A new globally complete monthly historical gridded mean sea level pressure dataset (HadSLP2): 1850-2004. *J Clim* 19:5816–5842. <https://doi.org/10.1175/JCLI3937.1>
- Andrews, T., Andrews, M. B., Bodas-Salcedo, A., Jones, G. S., Kuhlbrodt, T., Manners, J., Menary, M. B., Ridley, J., Ringer, M. A., Sellar, A. A., Senior, C. A., & Tang, Y. (2019). Forcings, Feedbacks, and Climate Sensitivity in HadGEM3-GC3.1 and UKESM1. *Journal of Advances in Modeling Earth Systems*, 11(12), 4377–4394.
- Bar-Joseph, Z., Gifford, D. K., & Jaakkola, T. S. (2001). Fast optimal leaf ordering for hierarchical clustering. *Bioinformatics*, 17(SUPPL. 1). https://doi.org/10.1093/bioinformatics/17.suppl_1.S22
- Bittner, M., Schmidt, H., Timmreck, C., & Sienz, F. (2016). Using a large ensemble of simulations to assess the Northern Hemisphere stratospheric dynamical response to tropical volcanic eruptions and its uncertainty. *Geophysical Research Letters*, 43(17), 9324-9332.
- Bladé I, Fortuny D, Van Oldenborgh GJ, Liebmann B (2012) The summer North Atlantic Oscillation in CMIP3 models and related uncertainties in projected summer drying in Europe. *J Geophys Res Atmos* 117:. <https://doi.org/10.1029/2012JD017816>
- Boé, J., Somot, S., Corre, L., & Nabat, P. (2020). Large discrepancies in summer climate change over Europe as projected by global and regional climate models: causes and consequences. *Climate Dynamics*, 54(5), 2981-3002.
- Browning, K. A. (2004) 'The sting at the end of the tail: Damaging winds associated with extratropical cyclones', *Quarterly Journal of the Royal Meteorological Society*. John Wiley & Sons, Ltd, 130(597), pp. 375–399.

- Brunner, L., Hauser, M., Lorenz, R., & Beyerle, U. (2020c). The ETH Zurich CMIP6 next generation archive: technical documentation. 10.
- Brunner, L., Lorenz, R., Zumwald, M., & Knutti, R. (2019). Quantifying uncertainty in European climate projections using combined performance-independence weighting. *Environmental Research Letters*, 14(12).
- Brunner, L., McSweeney, C., Ballinger, A. P., Befort, D. J., Benassi, M., Booth, B., Coppola, E., de Vries, H., Harris, G., Hegerl, G. C., Knutti, R., Lenderink, G., Lowe, J., Nogherotto, R., O'Reilly, C., Qasmi, S., Ribes, A., Stocchi, P., & Undorf, S. (2020a). Comparing Methods to Constrain Future European Climate Projections Using a Consistent Framework, *Journal of Climate*, 33(20), 8671-8692.
- Brunner, L., Pendergrass, A., Lehner, F., Merrifield, A., Lorenz, R., & Knutti, R. (2020b). Reduced global warming from CMIP6 projections when weighting models by performance and independence. *Earth System Dynamics Discussions*, 1–23.
- DAVINI, P. and D'ANDREA, F. (2020) 'From CMIP3 to CMIP6: Northern hemisphere atmospheric blocking simulation in present and future climate', *Journal of Climate*. Boston MA, USA: American Meteorological Society, 33(23), pp. 10021–10038. doi: 10.1175/JCLI-D-19-0862.1
- Deser, C., Phillips, A., Bourdette, V., & Teng, H. (2012). Uncertainty in climate change projections: the role of internal variability. *Climate dynamics*, 38(3), 527-546.
- Dong, B. et al. (2013) 'Variability of the North Atlantic summer storm track: mechanisms and impacts on European climate', *Environmental Research Letters*, 8(3), p. 34037. doi: 10.1088/1748-9326/8/3/034037
- Friederichs, P. and Hense, A. (2003) 'Statistical Inference in Canonical Correlation Analyses Exemplified by the Influence of North Atlantic SST on European Climate', *Journal of Climate*. Boston MA, USA: American Meteorological Society, 16(3), pp. 522–534. doi: 10.1175/1520-0442(2003)016<0522:SIICCA>2.0.CO;2.
- Hannachi, A., Jolliffe, I. T., & Stephenson, D. B. (2007). Empirical orthogonal functions and related techniques in atmospheric science: A review. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 27(9), 1119-1152.
- Hazeleger and co-authors, 2012. EC-Earth V2.2: description and validation of a new seamless earth system prediction model. *Climate Dynamics*, 39 (11), pp. 2611-2629.
- IPCC (2007) *Climate Change 2007: The Physical Science Basis*. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Edited by S. Solomon et al. Cambridge, UK and New York, NY, USA: Cambridge University Press. Available at: <https://www.ipcc.ch/report/ar4/wg1/>.
- IPCC (2013) *Climate Change 2013: The Physical Science Basis*. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Edited by T. F. Stocker et al. Cambridge, UK and New York, NY, USA: Cambridge University Press. Available at: <https://www.ipcc.ch/report/ar5/wg1/>
- Karlsson, J., & Svensson, G. (2013). Consequences of poor representation of Arctic sea-ice albedo and cloud-radiation interactions in the CMIP5 model ensemble. *Geophysical Research Letters*, 40(16), 4374–4379.
- Kaspi, Y. and Schneider, T. (2013) 'The Role of Stationary Eddies in Shaping Midlatitude Storm Tracks', *Journal of the Atmospheric Sciences*. Boston MA, USA: American Meteorological Society, 70(8), pp. 2596–2613. doi: 10.1175/JAS-D-12-082.1.

- Knutti, R. (2010) 'The end of model democracy?', *Climatic Change*, 102(3), pp. 395–404. doi: 10.1007/s10584-010-9800-2.
- Knutti, R., Sedláček, J., Sanderson, B. M., Lorenz, R., Fischer, E. M., & Eyring, V. (2017). A climate model projection weighting scheme accounting for performance and interdependence. *Geophysical Research Letters*, 44(4), 1909–1918.
- Liu, C., Allan, R. P., & Huffman, G. J. (2012). Co-variation of temperature and precipitation in CMIP5 models and satellite observations. *Geophysical Research Letters*, 39(13), 1–8.
- Lutz, A. F. et al. (2016) 'Selecting representative climate models for climate change impact studies: an advanced envelope-based selection approach', *International Journal of Climatology*. John Wiley & Sons, Ltd, 36(12), pp. 3988–4005. doi: <https://doi.org/10.1002/joc.4608>
- Maher, N., Matei, D., Milinski, S., & Marotzke, J. (2018). ENSO change in climate projections: Forced response or internal variability?. *Geophysical Research Letters*, 45(20), 11-390.
- Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Dobrynin, M., Kornblueh, L., ... & Marotzke, J. (2019). The Max Planck Institute Grand Ensemble: enabling the exploration of climate system variability. *Journal of Advances in Modeling Earth Systems*, 11(7), 2050-2069.
- McDermid, S. P. et al. (2014) 'The AgMIP Coordinated Climate-Crop Modeling Project (C3MP): Methods and Protocols', in *Handbook of Climate Change and Agroecosystems*. IMPERIAL COLLEGE PRESS (ICP Series on Climate Change Impacts, Adaptation, and Mitigation), pp. 191–220. doi: doi:10.1142/9781783265640_0008
- McSweeney, C. F. et al. (2015) 'Selecting CMIP5 GCMs for downscaling over multiple regions', *Climate Dynamics*, 44(11), pp. 3237–3260. doi: 10.1007/s00382-014-2418-8.
- McSweeney, C. F., Jones, R. G. and Booth, B. B. B. (2012) 'Selecting Ensemble Members to Provide Regional Climate Change Information', *Journal of Climate*. Boston MA, USA: American Meteorological Society, 25(20), pp. 7100–7121. doi: 10.1175/JCLI-D-11-00526.1
- McSweeney, J. Murphy, D. Sexton, J. Rostron, K. Yamazaki, G. Harris (2018): Selection of CMIP5 members to augment a perturbed-parameter ensemble of global realisations of future climate for the UKCP18 scenarios. Hadley Centre Technical Note: 102. https://digital.nmla.metoffice.gov.uk/IO_28f93601-3178-4c48-8621-fb4388ec66a1/
- Menary, M. B. et al. (2020) 'Aerosol-Forced AMOC Changes in CMIP6 Historical Simulations', *Geophysical Research Letters*, 47(14). doi: 10.1029/2020GL088166.
- Menary, M. B., Mignot, J. and Robson, J. (2021) 'Skilful decadal predictions of subpolar North Atlantic SSTs using CMIP model-analogues', *Environmental Research Letters*, 16(6), p. 64090. doi: 10.1088/1748-9326/ac06f
- Merrifield, A. L., Brunner, L., Lorenz, R., Medhaug, I., & Knutti, R. (2020). An investigation of weighting schemes suitable for incorporating large ensembles into multi-model ensembles. 807–834.
- Milinski, S., Maher, N., & Olonscheck, D. (2020). How large does a large ensemble need to be?. *Earth System Dynamics*, 11(4), 885-901.
- Müllner, D. (2011). Modern hierarchical, agglomerative clustering algorithms. 1973, 1–29. <http://arxiv.org/abs/1109.2378>
- Murphy JM, Harris GR, Sexton DMH, et al (2018) UKCP18 Land Projections: Science Report. 2018:1–191

- Overland, J. E. et al. (2011) 'Considerations in the Selection of Global Climate Models for Regional Climate Projections: The Arctic as a Case Study', *Journal of Climate*. Boston MA, USA: American Meteorological Society, 24(6), pp. 1583–1597. doi: 10.1175/2010JCLI3462.1.
- Philip and co-authors, 2020. Regional differentiation in climate change induced drought trends in the netherlands. *Environmental Research Letters*, 15, pp. 094081.
- Prein, A. F. et al. (2019) 'Simulating North American Weather Types With Regional Climate Models', *Frontiers in Environmental Science*, p. 36. Available at: <https://www.frontiersin.org/article/10.3389/fenvs.2019.00036>
- Righi, M., Andela, B., Eyring, V., Lauer, A., Predoi, V., Schlund, M., Vegas-Regidor, J., Bock, L., Brötz, B., De Mora, L., Diblen, F., Dreyer, L., Drost, N., Earnshaw, P., Hassler, B., Koldunov, N., Little, B., Loosveldt Tomas, S., & Zimmermann, K. (2020). Earth System Model Evaluation Tool (ESMValTool) v2.0-technical overview. Geoscientific Model Development.
- Ruane, A. C. et al. (2014) 'Carbon–Temperature–Water change analysis for peanut production under climate change: a prototype for the AgMIP Coordinated Climate-Crop Modeling Project (C3MP)', *Global Change Biology*. John Wiley & Sons, Ltd, 20(2), pp. 394–407. doi: <https://doi.org/10.1111/gcb.12412>.
- Ruane, A. C. and McDermid, S. P. (2017) 'Selection of a representative subset of global climate models that captures the profile of regional changes for integrated climate impacts assessment', *Earth Perspectives*, 4(1), p. 1. doi: 10.1186/s40322-017-0036-4.
- Sanderson, B. M., Knutti, R., & Caldwell, P. (2015). A representative democracy to reduce interdependency in a multimodel ensemble. *Journal of Climate*, 28(13), 5171–5194.
- Sanderson, B. M., Wehner, M., & Knutti, R. (2017). Skill and independence weighting for multi-model assessments. *Geoscientific Model Development*, 10(6), 2379–2395.
- Shepherd and co-authors, 2018. Storylines: an alternative approach to representing uncertainty in physical aspects of climate change. *Climatic change*, 151 (3-4), pp.555-571.
- Sillmann and co-authors, 2019. Physical modeling supporting a storyline approach. Tech. Rep. 2019:01, CICERO Senter for klimaforskning, <http://dx.doi.org/10.7892/boris.138767>.
- Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W., & Bronaugh, D. (2013). Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *Journal of Geophysical Research Atmospheres*, 118(4), 1716–1733.
- Sutton, R. T. and Hodson, D. L. R. (2003) 'Influence of the Ocean on North Atlantic Climate Variability 1871-1999', *Journal of Climate*. Boston MA, USA: American Meteorological Society, 16(20), pp. 3296–3313. doi: 10.1175/1520-0442(2003)016<3296:IOTOON>2.0.CO;2.
- Tebaldi, C., Debeire, K., Eyring, V., Fischer, E., & Fyfe, J. (2020). Climate model projections from the Scenario Model Intercomparison Project (ScenarioMIP) of CMIP6.
- Thompson, D. W., & Wallace, J. M. (1998). The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophysical research letters*, 25(9), 1297-1300.
- Van der Wiel and co-authors, 2019. Added value of large ensemble simulations for assessing extreme river discharge in a 2 °C warmer world. *Geophysical Research Letters*, 46 (4), pp.2093-2102.
- Van Oldenborgh and co-authors, 2021. Pathways and pitfalls in extreme event attribution. *Climatic Change*, 166 (1), pp.1-27.
- Wallace, J. M., & Gutzler, D. S. (1981). Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Monthly weather review*, 109(4), 784-812.

- Whetton, P. et al. (2007) 'Assessment of the use of current climate patterns to evaluate regional enhanced greenhouse response patterns of climate models', *Geophysical Research Letters*. John Wiley & Sons, Ltd, 34(14).
- White, J. W. et al. (2011) 'Methodologies for simulating impacts of climate change on crop production', *Field Crops Research*, 124(3), pp. 357–368.
- Wilcke, R. A. I., Barring, L. (2016). Selecting regional climate scenarios for impact modelling studies. *Environmental Modelling & Software*, 78, 191-201.
- Wills, R. C., Battisti, D. S., Hartmann, D. L., & Schneider, T. (2017). Extracting modes of variability and change from climate model ensembles. In *Proc. Seventh Int. Workshop on Climate Informatics: CI 2017* (pp. 25-28). NCAR Tech. Note NCAR/TN-5361PROC.
- Woollings T, Hannachi A, Hoskins B (2010) Variability of the North Atlantic eddy-driven jet stream. *Q J R Meteorol Soc* 136:856–868. <https://doi.org/10.1002/qj.625>
- Zappa and Shepherd, 2017. Storylines of atmospheric circulation change for European regional climate impact assessment. *Journal of Climate*, 30(16), 6561-6577.
- Zscheischler and Fischer, 2020. The record-breaking compound hot and dry 2018 growing season in Germany. *Weather and Climate Extremes*, 29, pp. 100270.

Appendix

A1 - Appendix to Science part 1 - Pruning the CMIP5/6 Family Tree using ClimWIP (ETHZ)

A1.1 INTRODUCTION -- This appendix serves as a supplement to Section 3.2 Science part 1 to explain the ClimWIP model sub-selection protocol in greater detail. The objective of the ClimWIP sub-selection protocol is to provide a recommendation of five climate models from the CMIP5 and CMIP6 archives that are best suited for applications requiring realistic estimates of regional European temperature and precipitation change. In this appendix, we present an example case and select models best suited for projecting mid-century Central European summer climate, surface air temperature (SAT) and precipitation (PR), change. Information about the CMIP5 and CMIP6 archives we select from is given in Section A1.2. The ClimWIP weighting method is described in Section A1.3, with predictor choice considerations detailed in Section A1.4. Finally, CMIP5/6 family tree construction is expanded on in Section A1.5 and the method to select independent high performing models is explained in Section A1.6.

A1.2 CMIP5/6 ARCHIVES -- We begin with a full CMIP5/6 ensemble comprised of all models (and all initial condition/perturbed physics ensemble members therein) with historical simulations and either RCP8.5 for CMIP5 model projections (Taylor et al., 2012) or Shared Socioeconomic Pathway 585 (ssp585) for CMIP6 model projections (Eyring et al., 2016; O'Neill et al., 2016). The CMIP5 historical period is extended from 2005 to 2014 using its RCP8.5 projection to coincide with the CMIP6 historical period. RCP8.5 and ssp585 projections anticipate an additional 8.5 W/m^2 of radiative forcing above pre-industrial levels by the end of the century. The scenarios represent futures in which anthropogenic emissions continue largely unmitigated by policy. Recent studies have found that current emissions are, unfortunately, on track with this scenario (Schwalm et al., 2020). It is important to note that while RCP8.5 and ssp585 projections experience similar aggregate radiative forcing by 2100, they differ in terms of the effective radiative forcing of their constituents (e.g., greenhouse gases and aerosols; Tebaldi et al., 2020). Because the scenarios are comparable and because ClimWIP was designed to compare behavior across generations of model development (Knutti et al., 2017), models are selected from a combined CMIP5/6 ensemble. At present, there are 78 CMIP5 and 198 CMIP6 simulations, listed in Tables A1a and A1b, that provide all fields necessary for metrics of performance (for Central European summer climate applications) and independence: SAT, PR, sea level pressure (SLP), sea surface temperature (SST), and the difference between all and clear sky downwelling shortwave radiation at the surface or shortwave cloud radiative effect (SWCRE). All fields are remapped onto a $2.5^\circ \times 2.5^\circ$ latitude–longitude grid. Additional CMIP6 models and members will be added to subsequent publications as fields become available in the CMIP6 next generation archive, a standardized repository used by researchers at ETH Zurich and made available upon request (Brunner, Hauser, et al., 2020).

<u>ACCESS1-0-r11p1 (5.16)</u>	CSIRO-Mk3-6-0-r5i1p1	<u>GISS-E2-H-r2i1p3 (6.86)</u>	<u>MIROC-ESM-CHEM-r11p1 (7.84)</u>
<u>ACCESS1-3-r11p1 (5.56)</u>	CSIRO-Mk3-6-0-r6i1p1	<u>GISS-E2-R-CC-r11p1 (6.48)</u>	<u>MIROC-ESM-r11p1 (7.46)</u>
CCSM4-r11p1	CSIRO-Mk3-6-0-r7i1p1	GISS-E2-R-r11p1	<u>MIROC5-r11p1 (5.45)</u>
CCSM4-r2i1p1	CSIRO-Mk3-6-0-r8i1p1	GISS-E2-R-r11p2	MIROC5-r2i1p1
<u>CCSM4-r3i1p1 (5.81)</u>	CSIRO-Mk3-6-0-r9i1p1	GISS-E2-R-r11p3	MIROC5-r3i1p1
CCSM4-r4i1p1	<u>CSIRO-Mk3-6-0-r10i1p1</u>	<u>GISS-E2-R-r2i1p1 (6.56)</u>	<u>MPI-ESM-LR-r11p1 (6.17)</u>
CCSM4-r5i1p1	CanESM2-r11p1	GISS-E2-R-r2i1p3	MPI-ESM-LR-r2i1p1
CCSM4-r6i1p1	<u>CanESM2-r2i1p1 (5.81)</u>	<u>HadGEM2-A0-r11p1 (5.67)</u>	MPI-ESM-LR-r3i1p1
<u>CESM1-BGC-r11p1 (5.96)</u>	CanESM2-r3i1p1	<u>HadGEM2-CC-r11p1 (6.39)</u>	<u>MPI-ESM-MR-r11p1 (6.27)</u>
CESM1-CAM5-r11p1	CanESM2-r4i1p1	HadGEM2-ES-r11p1	<u>MRI-CGCM3-r11p1 (7.19)</u>
<u>CESM1-CAM5-r2i1p1 (4.74)</u>	CanESM2-r5i1p1	HadGEM2-ES-r2i1p1	<u>MRI-ESM1-r11p1 (6.88)</u>
CESM1-CAM5-r3i1p1	FI0-ESM-r11p1	<u>HadGEM2-ES-r3i1p1 (5.54)</u>	<u>NorESM1-M-r11p1 (4.76)</u>
CNRM-CM5-r11p1	FI0-ESM-r2i1p1	HadGEM2-ES-r4i1p1	<u>NorESM1-ME-r11p1 (4.55)</u>
CNRM-CM5-r2i1p1	<u>FI0-ESM-r3i1p1 (6.43)</u>	IPSL-CM5A-LR-r11p1	<u>bcc-csm1-1-m-r11p1 (6.00)</u>
<u>CNRM-CM5-r4i1p1 (9.42)</u>	<u>GFDL-CM3-r11p1 (4.17)</u>	IPSL-CM5A-LR-r2i1p1	<u>bcc-csm1-1-r11p1 (6.68)</u>
CNRM-CM5-r6i1p1	<u>GFDL-ESM2G-r11p1 (5.73)</u>	(11.01)	<u>inmcm4-r11p1 (7.57)</u>
CNRM-CM5-r10i1p1	<u>GFDL-ESM2M-r11p1 (5.09)</u>	IPSL-CM5A-LR-r3i1p1	
<u>CSIRO-Mk3-6-0-r11p1 (6.58)</u>	<u>GISS-E2-H-CC-r11p1 (6.44)</u>	IPSL-CM5A-LR-r4i1p1	
CSIRO-Mk3-6-0-r2i1p1	GISS-E2-H-r11p1	<u>IPSL-CM5A-MR-r11p1 (8.39)</u>	
CSIRO-Mk3-6-0-r3i1p1	GISS-E2-H-r11p2	IPSL-CM5B-LR-r11p1	
CSIRO-Mk3-6-0-r4i1p1	GISS-E2-H-r11p3	(11.17)	
	GISS-E2-H-r2i1p1		

Table A1a. CMIP5 Members used, grouped by model ensemble. The highest performer for mid-century Central European summer climate applications in each ensemble is highlighted in red with its aggregated distance from observed, D_i , listed in parentheses. Smaller values of D_i indicate higher performance.

A1.3 ClimWIP -- Upon compiling the CMIP5/6 ensemble, we employ the ClimWIP method of model weighting to determine the performance and independence of each member listed in Tables A1a and A1b. ClimWIP is the performance and independence weighting scheme first applied in Knutti et al. (2017), which has brought together model evaluation and notions about commonalities in model development to provide justifiable representations of uncertainty in multi-model ensemble projections of future climate (Sanderson et al., 2015, 2017). It has since been integrated into the Earth System Model Evaluation Tool (ESMValTool; Righi et al., 2020) for general open use (https://docs.esmvaltool.org/en/latest/recipes/recipe_climwip.html).

Studies involving ClimWIP demonstrate that models that differ significantly from observations in relevant ways often exhibit unrealistic future behavior, increasing projection uncertainty (Knutti et al., 2017; Lorenz et al., 2018; Eyring et al., 2019). Additionally, subsequent work, carried out for WP2, highlights how replicate models or multiple members of the same model, two instances prevalent in the CMIP5/6 archives, can also affect estimates of multi-model ensemble uncertainty (Merrifield et al., 2020). Some of these so-called dependencies (e.g. multiple initial condition members of a model) are clear cut, but identifying others (e.g. code sharing) requires an insider's knowledge of climate model development (Knutti et al., 2013).

ACCESS-CM2-r11p1f1	CanESM5-r61p1f1	HadGEM3-GC31-LL-r21p1f3	MIROC6-r291p1f1
ACCESS-CM2-r21p1f1	CanESM5-r61p2f1	(5.17)	MIROC6-r301p1f1
ACCESS-CM2-r31p1f1 (5.30)	CanESM5-r71p1f1	HadGEM3-GC31-LL-r31p1f3	MIROC6-r311p1f1
ACCESS-ESM1-5-r11p1f1	CanESM5-r71p2f1	HadGEM3-GC31-LL-r41p1f3	MIROC6-r321p1f1
ACCESS-ESM1-5-r21p1f1	CanESM5-r81p1f1	HadGEM3-GC31-MM-r11p1f3	MIROC6-r331p1f1
ACCESS-ESM1-5-r31p1f1	CanESM5-r81p2f1	HadGEM3-GC31-MM-r21p1f3	MIROC6-r341p1f1
ACCESS-ESM1-5-r41p1f1	CanESM5-r91p1f1	HadGEM3-GC31-MM-r31p1f3	MIROC6-r351p1f1
ACCESS-ESM1-5-r51p1f1	CanESM5-r91p2f1	HadGEM3-GC31-MM-r41p1f3	MIROC6-r361p1f1
ACCESS-ESM1-5-r61p1f1	CanESM5-r101p1f1	(4.39)	MIROC6-r371p1f1
ACCESS-ESM1-5-r71p1f1	CanESM5-r101p2f1	INM-CM4-8-r11p1f1 (8.57)	MIROC6-r381p1f1
ACCESS-ESM1-5-r81p1f1	CanESM5-r111p1f1	INM-CM5-0-r11p1f1 (7.51)	MIROC6-r391p1f1
ACCESS-ESM1-5-r91p1f1	CanESM5-r111p2f1	IPSL-CM6A-LR-r11p1f1	MIROC6-r401p1f1
ACCESS-ESM1-5-r101p1f1 (5.31)	CanESM5-r121p1f1 (5.42)	IPSL-CM6A-LR-r21p1f1	MIROC6-r411p1f1
AWI-CM-1-1-MR-r11p1f1 (4.07)	CanESM5-r121p2f1	IPSL-CM6A-LR-r31p1f1	MIROC6-r421p1f1
CAS-ESM2-0-r11p1f1	CanESM5-r131p1f1	IPSL-CM6A-LR-r31p1f1 (5.10)	MIROC6-r431p1f1
CAS-ESM2-0-r31p1f1 (8.22)	CanESM5-r131p2f1	IPSL-CM6A-LR-r41p1f1	MIROC6-r441p1f1 (6.09)
CESM2-WACCM-r11p1f1 (4.86)	CanESM5-r141p1f1	IPSL-CM6A-LR-r61p1f1	MIROC6-r451p1f1
CESM2-WACCM-r21p1f1	CanESM5-r141p2f1	IPSL-CM6A-LR-r141p1f1	MIROC6-r461p1f1
CESM2-WACCM-r31p1f1	CanESM5-r151p1f1	MIROC-ES2L-r11p1f2 (5.75)	MIROC6-r471p1f1
CESM2-r11p1f1	CanESM5-r161p1f1	MIROC-ES2L-r21p1f2	MIROC6-r481p1f1
CESM2-r21p1f1	CanESM5-r161p2f1	MIROC-ES2L-r31p1f2	MIROC6-r491p1f1
CESM2-r41p1f1	CanESM5-r171p1f1	MIROC-ES2L-r51p1f2	MIROC6-r501p1f1
CESM2-r101p1f1 (4.97)	CanESM5-r171p2f1	MIROC-ES2L-r61p1f2	MPI-ESM1-2-HR-r11p1f1
CESM2-r111p1f1	CanESM5-r181p1f1	MIROC-ES2L-r71p1f2	MPI-ESM1-2-HR-r21p1f1 (4.47)
CMCC-CM2-SR5-r11p1f1 (4.79)	CanESM5-r181p2f1	MIROC-ES2L-r91p1f2	MPI-ESM1-2-LR-r11p1f1
CMCC-ESM2-r11p1f1 (4.51)	CanESM5-r191p1f1	MIROC-ES2L-r101p1f2	MPI-ESM1-2-LR-r21p1f1
CNRM-CM6-1-HR-r11p1f2 (5.76)	CanESM5-r191p2f1	MIROC6-r11p1f1	MPI-ESM1-2-LR-r31p1f1
CNRM-CM6-1-r11p1f2	CanESM5-r201p1f1	MIROC6-r21p1f1	MPI-ESM1-2-LR-r41p1f1
CNRM-CM6-1-r21p1f2	CanESM5-r201p2f1	MIROC6-r31p1f1	MPI-ESM1-2-LR-r51p1f1
CNRM-CM6-1-r31p1f2	CanESM5-r211p1f1	MIROC6-r41p1f1	MPI-ESM1-2-LR-r61p1f1
CNRM-CM6-1-r41p1f2	CanESM5-r211p2f1	MIROC6-r51p1f1	MPI-ESM1-2-LR-r71p1f1
CNRM-CM6-1-r51p1f2	CanESM5-r221p1f1	MIROC6-r61p1f1	MPI-ESM1-2-LR-r81p1f1
CNRM-CM6-1-r61p1f2 (4.95)	CanESM5-r221p2f1	MIROC6-r71p1f1	MPI-ESM1-2-LR-r91p1f1
CNRM-ESM2-1-r11p1f2 (4.85)	CanESM5-r231p1f1	MIROC6-r81p1f1	MPI-ESM1-2-LR-r101p1f1 (5.28)
CNRM-ESM2-1-r21p1f2	CanESM5-r231p2f1	MIROC6-r91p1f1	MRI-ESM2-0-r11p1f1 (4.27)
CNRM-ESM2-1-r31p1f2	CanESM5-r241p1f1	MIROC6-r101p1f1	MRI-ESM2-0-r11p1f1
CNRM-ESM2-1-r41p1f2	CanESM5-r241p2f1	MIROC6-r111p1f1	NESM3-r11p1f1 (6.27)
CNRM-ESM2-1-r51p1f2	CanESM5-r251p1f1	MIROC6-r121p1f1	NESM3-r21p1f1
CanESM5-CanOE-r11p2f1	CanESM5-r251p2f1	MIROC6-r131p1f1	NoRESM2-MM-r11p1f1 (4.65)
CanESM5-CanOE-r21p2f1 (5.64)	E3SM-1-1-r11p1f1 (6.15)	MIROC6-r141p1f1	TaiESM1-r11p1f1 (4.50)
CanESM5-CanOE-r31p2f1	FGOALS-f3-L-r11p1f1 (7.08)	MIROC6-r151p1f1	UKESM1-0-LL-r11p1f2
CanESM5-r11p1f1	FGOALS-g3-L-r11p1f1 (6.62)	MIROC6-r161p1f1	UKESM1-0-LL-r21p1f2
CanESM5-r11p2f1	FGOALS-g3-r21p1f1	MIROC6-r171p1f1	UKESM1-0-LL-r31p1f2
CanESM5-r21p1f1	FI0-ESM-2-0-r11p1f1	MIROC6-r181p1f1	UKESM1-0-LL-r41p1f2 (5.29)
CanESM5-r21p2f1	FI0-ESM-2-0-r21p1f1	MIROC6-r191p1f1	UKESM1-0-LL-r81p1f2
CanESM5-r31p1f1	FI0-ESM-2-0-r31p1f1 (6.43)	MIROC6-r201p1f1	
CanESM5-r31p2f1	GFDL-CM4-r11p1f1 (4.50)	MIROC6-r211p1f1	
CanESM5-r41p1f1	GFDL-ESM4-r11p1f1 (4.49)	MIROC6-r221p1f1	
CanESM5-r41p2f1	GISS-E2-1-G-r11p3f1 (4.94)	MIROC6-r231p1f1	
CanESM5-r51p1f1	HadGEM3-GC31-LL-r11p1f3	MIROC6-r241p1f1	
CanESM5-r51p2f1		MIROC6-r251p1f1	
		MIROC6-r261p1f1	
		MIROC6-r271p1f1	
		MIROC6-r281p1f1	

Table A1b. CMIP6 Members used, grouped by model ensemble. The highest performer for mid-century Central European summer climate applications in each ensemble is highlighted in red with its aggregated distance from observed, D_i , listed in parentheses. Smaller values of D_i indicate higher performance.

To handle the issues associated with model performance and dependence, a straightforward Gaussian-based weighting function was proposed by Knutti et al., 2017. In the numerator, the function weights a member of a multi-model ensemble according to its distance from observations (over user-selected relevant climate metrics). In the denominator, the member's weight is scaled according to how similar it is to other models (to account for known and unknown dependencies):

$$W_i = \frac{e^{-\frac{D_i^2}{\sigma_D^2}}}{1 + \sum_{j \neq i}^M e^{-\frac{S_{ij}^2}{\sigma_S^2}}}$$

D_i defines the aggregate distance from observations of an ensemble member i over a set of performance metrics. S_{ij} defines the distance between ensemble member i and all other members of the ensemble over a separate set of dependence metrics. Also required are two shape parameters, σ_D to determine how strongly a model will be penalized for lacking historical realism and σ_S to determine how similar two models must be to be considered replicates. The shape parameters are set by user discretion; σ_D can be estimated through model-based calibration (Brunner et al., 2019), while σ_S should be set between the largest inter-member (of the same model) S_{ij} and the smallest inter-model S_{ij} as

described in Merrifield et al. (2020). Because the shape parameters add additional degrees of freedom in the determination of performance and independence, we simply use D_i and S_{ij} to select high performing, independent ensemble members rather than the Gaussian-based weighting functions.

In the first selection round, D_i is used to select the member deemed highest performing or most suitable member for projecting Central European summer climate in each model ensemble within the CMIP5/6 archives to create informed “one model, one vote” ensembles. We define model ensemble as an initial condition and/or perturbed physics ensemble where members share the same model name (e.g. CESM1-CAM5 or ACCESS-ESM1-5) . Model ensembles are grouped in Tables A1a,b with the highest performer highlighted in red and its D_i given in parentheses. The lower the value of D_i , the closer a model is to observed and the more suitable it is for Central European summer climate applications.

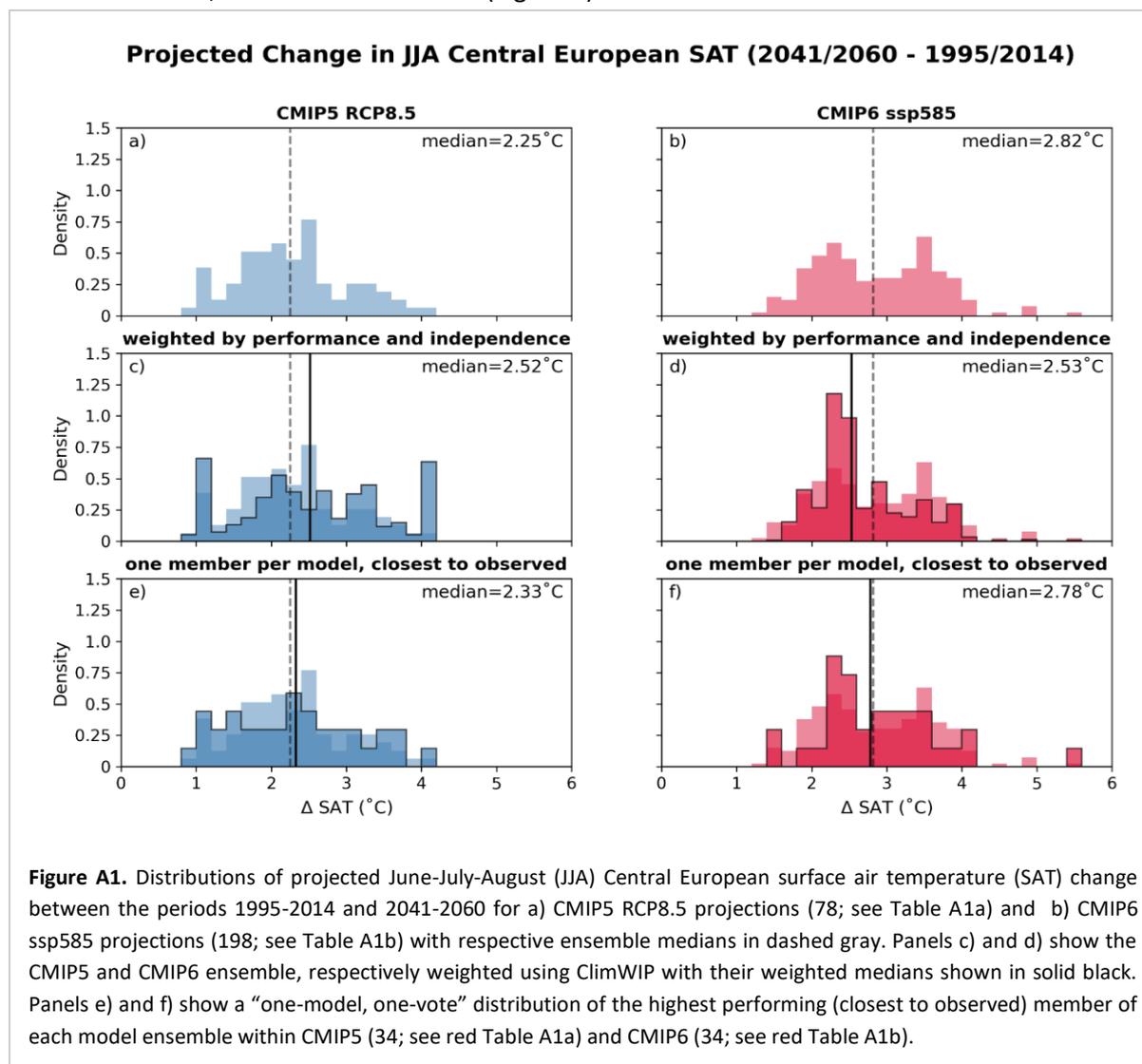
A1.4 PERFORMANCE AND INDEPENDENCE METRICS -- The two keys to the ClimWIP sub-selection protocol are to determine if a model is suited to the given task, simulating Central European summer climate, and to determine whether a model is suitably distinct from its counterparts. It is important to note that no one model is better than all others at simulating every aspect of the climate system; all models have strengths and weaknesses. Models deemed high performers in this example may not be the best suited to a different region, season, or objective. It is also important to recognize that most climate models are not truly independent from their peers; many share a component model (Boé, 2018), are based on the same predecessor (Masson & Knutti, 2011), or are variants on an institution’s base model (Leduc et al., 2016). An awareness of these “model families” is required to avoid choosing multiple versions of the same model with different names and thus losing the advantages of a multi-model ensemble (Abramowitz et al., 2019).

Within the scope of EUCP’s focus on projected European climate change, we target realistic change in June-July-August (JJA) average SAT (Fig. A1) and PR (Fig. A2) between 1995-2014 and 2041-2060 mean states (mid-century) in the Central European SREX region (CEU; Iturbide et al., 2020). In Figs. A1 and A2, distributions of mid-century JJA CEU SAT and PR change are shown for:

- The full (a) CMIP5 and (b) CMIP6 ensembles listed in Tables A1a,b respectively,
- the same (c) CMIP5 and (d) CMIP6 ensembles weighted using ClimWIP
- one-model, one-vote distributions of (e) CMIP5 and (f) CMIP6 using only the highest performing (closest to observed) member of each model ensemble within CMIP5 and CMIP6 (indicated in red in Tables A1a,b).

In each panel, the full ensemble median (dashed gray line) serves as an indicator of how weighting and model sub-selection introduces distributional shifts. For JJA CEU SAT change, the full CMIP6 ensemble has a higher median warming by mid-century than the CMIP5 ensemble, largely due to its bimodal distribution containing models with higher climate sensitivity (more warming in response to doubling atmospheric CO_2) than previous generations (Flynn & Mauritsen, 2020; Meehl et al., 2020; Zelinka et al., 2020). Weighting the full JJA CEU SAT change distributions using ClimWIP reconciles median warming in CMIP5 and CMIP6 (Fig. A1c,d; solid black line) to around 2.5°C , though the weighting does amplify the influence of two CMIP5 high performers (GFDL-CM3-r1i1p1 and GFDL-ESM2M-r1i1p1) with mid-century warming estimates at the upper and lower ends of the distribution. The one-model,

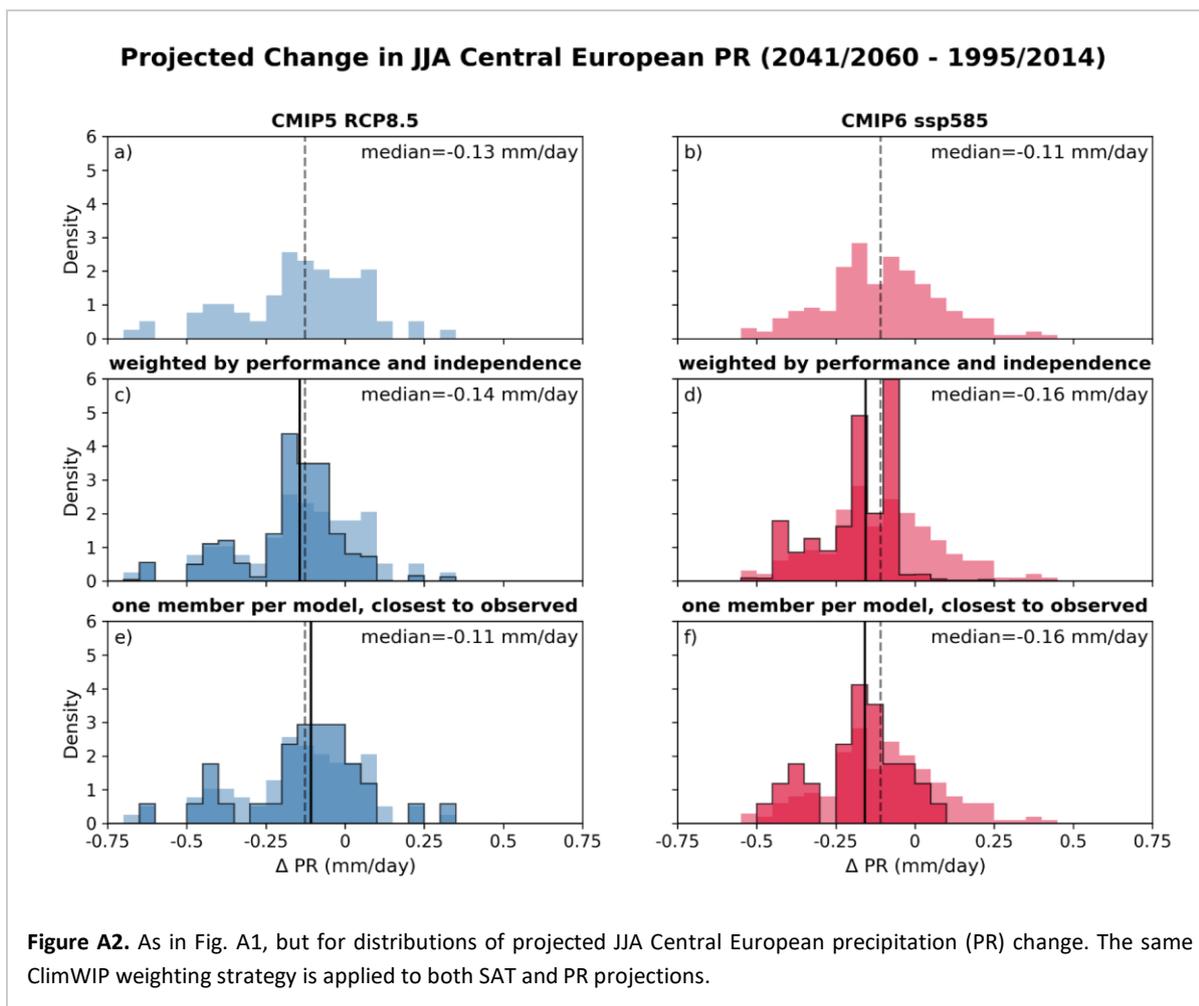
one vote JJA CEU SAT change distributions more closely resemble their full ensemble counterparts in median value (Fig. A1e,f; solid black line), but sub-selection notably reduces the higher warming peak of the bimodal full CMIP6 distribution. For JJA CEU PR change, both the ClimWIP weighting and model sub-selection narrow the spread of the full CMIP5 and CMIP6 distributions towards a median reduction of between -0.1 and -0.2 mm/day or around 5-7% of the JJA CEU area average (not shown). In the case of CMIP6, the ClimWIP weighting decreases the likelihood of a slight increase in future precipitation in the region by eliminating the considerable internal variability present within CMIP6's two 50-member model ensembles, CanESM5 and MIROC6 (Fig. A2d).



Knowledge of the plausible mechanisms underpinning future European temperature and precipitation change is required to aggregate a set of metrics to define model performance. For sub-selection, we use a mix of local and global climatological predictors we anticipate will form a base set from which performance metrics in other regions and seasons can be constructed:

- Annual average European SAT Climatology; 1950-1969
- Annual average European SAT Climatology; 1995-2014
- JJA average Central European Station PR Climatology; 1995-2014
- JJA average Central European SREX (CEU) SWCRE Climatology; 2001-2018
- Annual average North Atlantic SST Climatology; 1995-2014

- Annual average Southern Hemisphere midlatitude SWCRE Climatology; 2001-2018

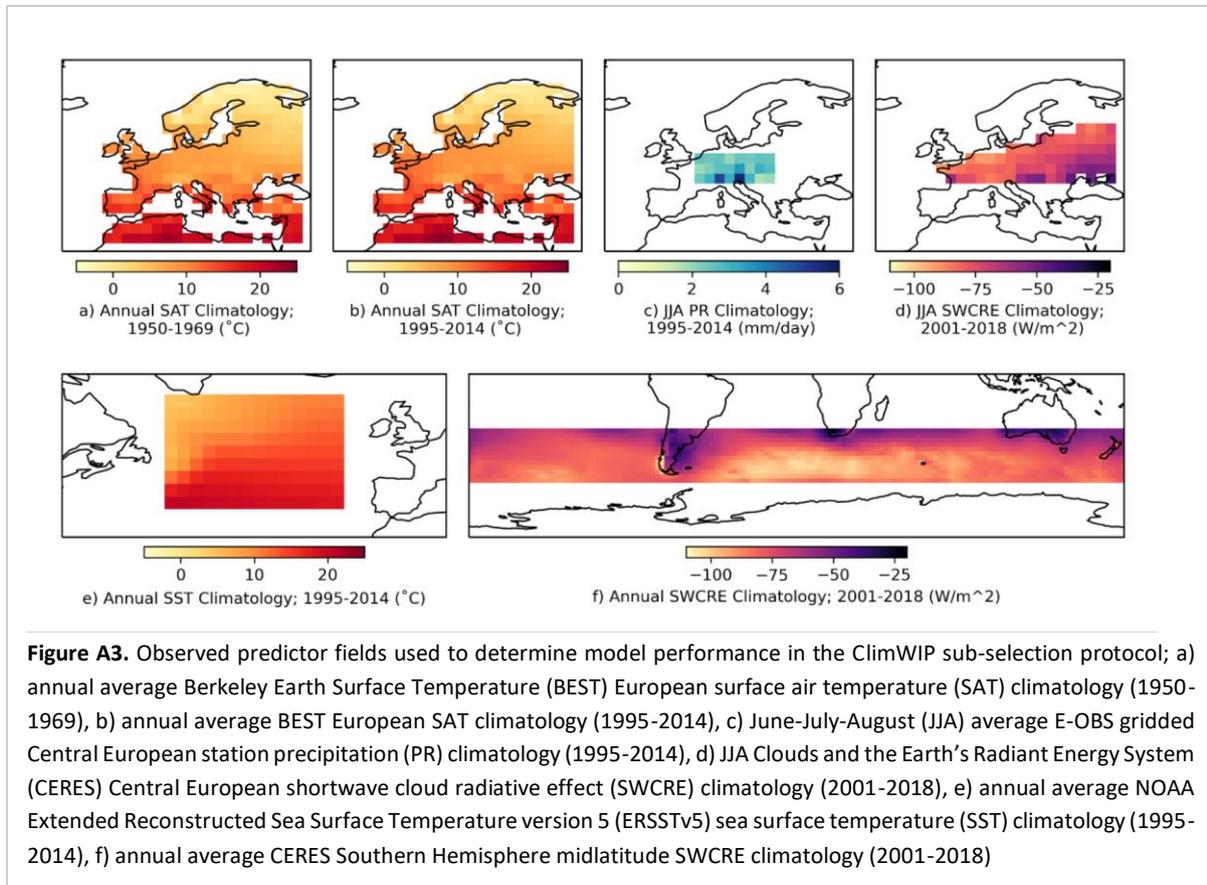


Local predictors have the potential to reveal major historical biases that erode confidence in future projections: for example, warming biases (perhaps due to biases in local cloud cover) may affect a model’s ability to warm a realistic amount in the future. Precipitation biases may lead to moisture availability and local land-atmosphere interaction issues. We use two periods of SAT climatology to avoid penalizing models for differing from observed by chance due to being in a different phase of internal variability (Deser et al., 2012). Regional-scale predictors, such as North Atlantic SST climatology, have also been linked to European summer climate variability through interactions with atmospheric circulation (Borchert et al., 2019). And finally, with the advent of CMIP6 and models with higher climate sensitivity, we incorporate a metric that affects how much a model warms globally (as European change is superposed on global change) into the performance predictor set. One such metric is SWCRE in low cloud regions of the Southern Hemisphere (Zelinka et al., 2020). Models that historically underestimate Southern Hemisphere low clouds decks do not have them present to counteract future radiative warming increases associated with the Hadley cell and its “high cloud curtain” moving poleward (Lipat et al., 2017; Tselioudis et al., 2016).

Model performance is benchmarked against the following observational estimates, shown in Fig. A3:

- SAT: Berkeley Earth Surface Temperature merged temperature (BEST; Rohde et al., 2013)

- SLP: NOAA-CIRES-DOE 20th Century Reanalysis V3 reanalysis (Bloomfield et al., 2018)
- PR: European-wide station data based E-OBS dataset (E-OBS; Cornes et al., 2018)
- SWCRE: Clouds and the Earth’s Radiant Energy System (CERES) Energy Balanced and Filled All- and Clear-Sky shortwave surface flux products (Loeb et al., 2018, 2020)
- SST: NOAA Extended Reconstructed Sea Surface Temperature version 5 (Huang et al., 2017)



ClimWIP also accommodates more than one observational dataset per field (to represent observational uncertainty), and as such, additional datasets may be included in the future publication to assess the sensitivity of the selection protocol to the choice of observational dataset(s). A cosine latitude weighted root-mean-square-error distance metric is computed between CMIP5/6 members and observations for the performance metrics and the metric distances are normalized by their mean value and added together to define the performance D_i . For the Central European summer climate predictor set listed above, D_i ranges from 4.1 to 11.3 with lower values indicating higher performance (see Tables A1a,b).

Determining model dependence requires a different strategy and thus different predictors than those used to define model performance. Large-scale, long term average values (model properties with minimal internal variability) are required to identify initial condition ensemble members as members of the same model and to assign degrees of dependence to closely related models. We find that the dependence predictor set proposed in Merrifield et al. (2020), annual global land temperature and

Northern Hemisphere SLP mean states over the period 1950-2014, effectively groups model ensembles within the CMIP5/6 ensemble and is suitable for use in family tree construction.

A1.5 CONSTRUCTING A CMIP5/6 FAMILY TREE -- The second step of the ClimWIP sub-selection protocol employs hierarchical cluster analysis to build a model dependence family tree dendrogram featuring the highest performing member of each model family (Bar-Joseph et al., 2001; Müllner, 2011). Euclidean distance, computed from the center of the dependence weight distribution (average-linkage criteria), is used to group members into model families. The clustering is done iteratively, linking the two closest models and so forth, with distances to existing clusters calculated using the average of their elements. More dependent models with many “close relatives” have more branches across the tree from left to right, while models with fewer similar counterparts have fewer branches.

A1.6 SELECTING INDEPENDENT HIGH PERFORMERS -- To recommend a set of independent high performers, a set of models can be selected which minimize a performance/independence cost function described in the main text:

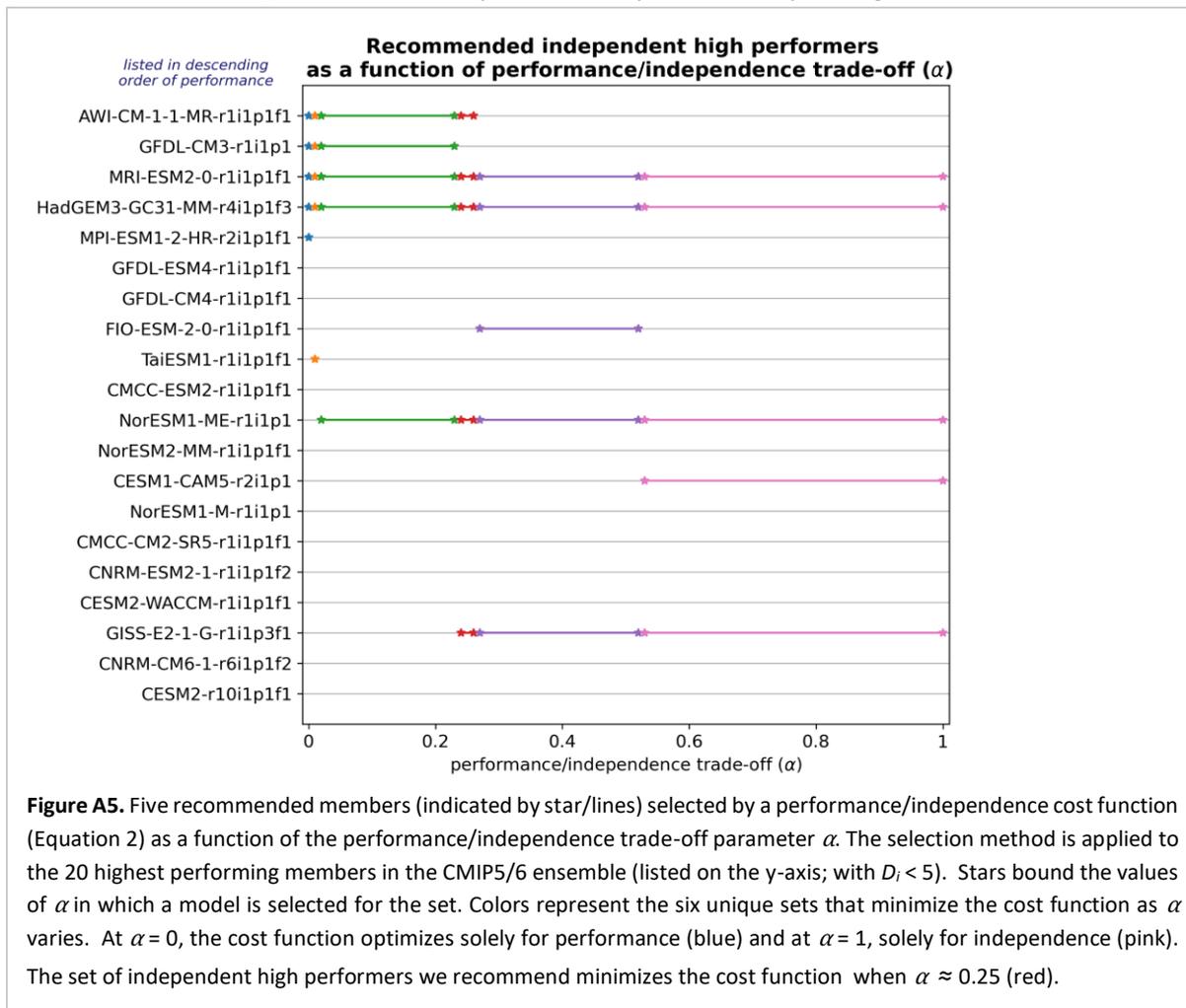
$$C_{\alpha}(m_1, \dots, m_5) = (1 - \alpha) \cdot \frac{1}{\min(D_i)} \sum_{i=1}^5 D_i - \alpha \cdot \frac{1}{2} \sum_{j \neq i}^5 \frac{S_{ij}}{S_{iN}}$$

$$S_{iN} = \frac{1}{N} \cdot \sum_{j \neq i}^N S_{ij}$$

In this instance, we select sets of 5 models, though sets of any size less than the total size of the selection pool can be chosen. To reduce computational expense, a performance threshold (a user choice) can be set to eliminate lower performing models from the selection pool. Because the cost function is computed for all possible 5 model combinations, a smaller selection pool produces significantly fewer combinations. In our case, the original 68 models have 10,424,128 possible 5 model combinations. We chose a performance threshold of $D_i < 5$ and the remaining 20 models have only 15,504 possible 5 model combinations. Additionally, the performance threshold can ensure that models which are very independent from all other models, for example due to extreme biases with respect to observations, do not dominate the cost function.

The performance term of the cost function is positive and smallest for smaller values of D_i , i.e. higher performing model combinations. By subtracting the independence term, the cost function gets progressively more negative as models differ from one another. The relative importance of these two terms is set by the performance/independence trade-off parameter α . The parameter allows a user to decide the importance of independence in model selection for their use case; Figure A5 shows the sensitivity of the cost function to choice of α . As α is varied from 0 to 1, six sets of 5 models, selected based on their minimization of C_{α} , arise (Fig. A5). For $\alpha = 0$, the cost function is based solely on performance and selects the 5 highest performing models (Fig. A5; blue): AWI-CM-1-1-MR-r1i1p1f1, GFDL-CM3-r1i1p1, MRI-ESM2-0-r1i1p1f1, HadGEM3-GC31-MM-r4i1p1f3, and MPI-ESM1-2-HR-r2i1p1f1. Two of these models, AWI-CM-1-1-MR and MPI-ESM1-2-HR, are each other’s closest relative in the CMIP5/6 family tree (Fig. 1 main text; orange), and thus not independent. For α between 0 and 1, both the performance and independence terms contribute to the cost function. For most applications, we recommend the 5 independent high performers given by $\alpha \approx 0.25$ (Fig. A5; red): AWI-CM-1-1-MR-r1i1p1f1, MRI-ESM2-0-r1i1p1f1, HadGEM3-GC31-MM-r4i1p1f3, NorESM1-ME-r1i1p1,

and GISS-E2-1-G-r1i1p3f1. The set retains the highest performer, AWI-CM-1-1-MR, while selecting the remaining models from different model families within the CMIP5/6 family tree. For $\alpha = 1$, the cost function is based solely on independence and selects the 5 most independent models from the selection pool (of models with $D_i < 5$). Shown in pink in Fig. A5, the 5 most independent models in the set minimize the cost function starting at $\alpha = 0.53$: MRI-ESM2-0-r1i1p1f1, HadGEM3-GC31-MM-r4i1p1f3, NorESM1-ME-r1i1p1, CESM1-CAM5-r2i1p1, and GISS-E2-1-G-r1i1p3f1. NorESM1-ME and CESM1-CAM5 originate from subsequent generations of National Center of Atmospheric Research (NCAR) model development (Fig. 1 main text, tan), CCSM4 and CESM1 (Meehl et al., 2013). This highlights that model independence cannot be defined solely on the basis of a modelling center (Merrifield et al. 2020), as model development can represent a step-change in simulation outcome.



REFERENCES

- Abramowitz, G., Herger, N., Gutmann, E., Hammerling, D., Knutti, R., Leduc, M., Lorenz, R., Pincus, R., & Schmidt, G. A. (2019). ESD Reviews: Model dependence in multi-model climate ensembles: Weighting, sub-selection and out-of-sample testing. In *Earth System Dynamics* (Vol. 10, Issue 1, pp. 91–105). Copernicus GmbH. <https://doi.org/10.5194/esd-10-91-2019>
- Andrews, T., Andrews, M. B., Bodas-Salcedo, A., Jones, G. S., Kuhlbrodt, T., Manners, J., Menary, M. B., Ridley, J., Ringer, M. A., Sellar, A. A., Senior, C. A., & Tang, Y. (2019). Forcings, Feedbacks, and Climate Sensitivity in HadGEM3-GC3.1 and UKESM1. *Journal of Advances in Modeling Earth Systems*, 11(12), 4377–4394. <https://doi.org/10.1029/2019MS001866>

- Bar-Joseph, Z., Gifford, D. K., & Jaakkola, T. S. (2001). Fast optimal leaf ordering for hierarchical clustering. *Bioinformatics*, 17(SUPPL. 1). https://doi.org/10.1093/bioinformatics/17.suppl_1.S22
- Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevåg, A., Seland, Ø., Drange, H., Roelandt, C., Seierstad, I. A., Hoose, C., and Kristjánsson, J. E.: The Norwegian Earth System Model, NorESM1-M – Part 1: Description and basic evaluation of the physical climate, *Geosci. Model Dev.*, 6, 687–720, <https://doi.org/10.5194/gmd-6-687-2013>, 2013.
- Bloomfield, H. C., Shaffrey, L. C., Hodges, K. I., & Vidale, P. L. (2018). A critical assessment of the long-term changes in the wintertime surface Arctic Oscillation and Northern Hemisphere storminess in the ERA20C reanalysis. *Environmental Research Letters*. <https://doi.org/10.1088/1748-9326/aad5c5>
- Boé, J. (2018). Interdependency in Multimodel Climate Projections: Component Replication and Result Similarity. *Geophysical Research Letters*, 45(6), 2771–2779. <https://doi.org/10.1002/2017GL076829>
- Borchert, L. F., Pohlmann, H., Baehr, J., Neddermann, N. C., Suarez-Gutierrez, L., & Müller, W. A. (2019). Decadal Predictions of the Probability of Occurrence for Warm Summer Temperature Extremes. *Geophysical Research Letters*, 46(23), 14042–14051. <https://doi.org/10.1029/2019GL085385>
- Brunner, L., Hauser, M., Lorenz, R., & Beyerle, U. (2020b). *The ETH Zurich CMIP6 next generation archive: technical documentation*. 10. <https://doi.org/10.5281/zenodo.3734128>.Contact.
- Brunner, L., Lorenz, R., Zumwald, M., & Knutti, R. (2019). Quantifying uncertainty in European climate projections using combined performance-independence weighting. *Environmental Research Letters*, 14(12). <https://doi.org/10.1088/1748-9326/ab492f>
- Brunner, L., Pendergrass, A., Lehner, F., Merrifield, A., Lorenz, R., & Knutti, R. (2020a). Reduced global warming from CMIP6 projections when weighting models by performance and independence. *Earth System Dynamics Discussions*, 1–23. <https://doi.org/10.5194/esd-2020-23>
- Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., & Jones, P. D. (2018). An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets. *Journal of Geophysical Research: Atmospheres*, 123(17), 9391–9409. <https://doi.org/10.1029/2017JD028200>
- Deser, C., Phillips, A., Bourdette, V., & Teng, H. (2012). Uncertainty in climate change projections: The role of internal variability. *Climate Dynamics*, 38(3–4), 527–546. <https://doi.org/10.1007/s00382-010-0977-x>
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5), 1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>
- Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., Collins, W. D., Gier, B. K., Hall, A. D., Hoffman, F. M., Hurtt, G. C., Jahn, A., Jones, C. D., Klein, S. A., Krasting, J. P., Kwiatkowski, L., Lorenz, R., Maloney, E., Meehl, G. A., ... Williamson, M. S. (2019). Taking climate model evaluation to the next level. *Nature Climate Change*, 9(2), 102–110. <https://doi.org/10.1038/s41558-018-0355-y>
- Flynn, C. M., & Mauritsen, T. (2020). On the climate sensitivity and historical warming evolution in recent coupled model ensembles. *Atmospheric Chemistry and Physics*, 20(13), 7829–7842. <https://doi.org/10.5194/acp-20-7829-2020>
- Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., Menne, M. J., Smith, T. M., Vose, R. S., & Zhang, H. M. (2017). Extended reconstructed Sea surface temperature, Version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *Journal of Climate*. <https://doi.org/10.1175/JCLI-D-16-0836.1>
- Iturbide, M., Gutiérrez, J. M., Alves, L. M., Bedia, J., Cimadevilla, E., Cofiño, A., Cerezo-Mota, R., Di Luca, A., Faria, S. H., Gorodetskaya, I., Hauser, M., Herrera, S., Hewitt, H., Hennessy, K., Jones, R., Krakovska, S., Manzananas, R., Marínez-Castro, D., Narisma, G. T., ... Vera, C. (2020). An update of IPCC

climate reference regions for subcontinental analysis of climate model data: Definition and aggregated datasets. *Earth System Science Data Discussions*, April, 1–16.

<https://doi.org/10.5194/essd-2019-258>

- Karlsson, J., & Svensson, G. (2013). Consequences of poor representation of Arctic sea-ice albedo and cloud-radiation interactions in the CMIP5 model ensemble. *Geophysical Research Letters*, 40(16), 4374–4379. <https://doi.org/10.1002/grl.50768>
- Knutti, R., Masson, D., & Gettelman, A. (2013). Climate model genealogy: Generation CMIP5 and how we got there. *Geophysical Research Letters*, 40(6), 1194–1199. <https://doi.org/10.1002/grl.50256>
- Knutti, R., Sedláček, J., Sanderson, B. M., Lorenz, R., Fischer, E. M., & Eyring, V. (2017). A climate model projection weighting scheme accounting for performance and interdependence. *Geophysical Research Letters*, 44(4), 1909–1918. <https://doi.org/10.1002/2016GL072012>
- Leduc, M., Laprise, R., de Elía, R., & Šeparović, L. (2016). Is institutional democracy a good proxy for model independence? *Journal of Climate*, 29(23), 8301–8316. <https://doi.org/10.1175/JCLI-D-15-0761.1>
- Lipat, B. R., Tselioudis, G., Grise, K. M., & Polvani, L. M. (2017). CMIP5 models' shortwave cloud radiative response and climate sensitivity linked to the climatological Hadley cell extent. *Geophysical Research Letters*. <https://doi.org/10.1002/2017GL073151>
- Liu, C., Allan, R. P., & Huffman, G. J. (2012). Co-variation of temperature and precipitation in CMIP5 models and satellite observations. *Geophysical Research Letters*, 39(13), 1–8. <https://doi.org/10.1029/2012GL052093>
- Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., Liang, L., Mitrescu, C., Rose, F. G., & Kato, S. (2018). Clouds and the Earth'S Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) top-of-atmosphere (TOA) edition-4.0 data product. *Journal of Climate*, 31(2), 895–918. <https://doi.org/10.1175/JCLI-D-17-0208.1>
- Loeb, N. G., Rose, F. G., Kato, S., Rutan, D. A., Su, W., Wang, H., Doelling, D. R., Smith, W. L., & Gettelman, A. (2020). Toward a consistent definition between satellite and model clear-sky radiative fluxes. *Journal of Climate*, 33(1), 61–75. <https://doi.org/10.1175/JCLI-D-19-0381.1>
- Lorenz, R., Herger, N., Sedláček, J., Eyring, V., Fischer, E. M., & Knutti, R. (2018). Prospects and Caveats of Weighting Climate Models for Summer Maximum Temperature Projections Over North America. *Journal of Geophysical Research: Atmospheres*, 123(9), 4509–4526. <https://doi.org/10.1029/2017JD027992>
- Masson, D., & Knutti, R. (2011). Climate model genealogy. *Geophysical Research Letters*, 38(8), 1–4. <https://doi.org/10.1029/2011GL046864>
- Meehl, G. A., Senior, C. A., Eyring, V., Flato, G., Lamarque, J. F., Stouffer, R. J., Taylor, K. E., & Schlund, M. (2020). Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models. *Science Advances*, 6(26), 1–11. <https://doi.org/10.1126/sciadv.aba1981>
- Meehl, G. A., Washington, W. M., Arblaster, J. M., Hu, A., Teng, H., Kay, J. E., Gettelman, A., Lawrence, D. M., Sanderson, B. M., & Strand, W. G. (2013). Climate change projections in CESM1(CAM5) compared to CCSM4. *Journal of Climate*, 26(17), 6287–6308. <https://doi.org/10.1175/JCLI-D-12-00572.1>
- Merrifield, A. L., Brunner, L., Lorenz, R., Medhaug, I., & Knutti, R. (2020). *An investigation of weighting schemes suitable for incorporating large ensembles into multi-model ensembles*. 807–834.
- Müllner, D. (2011). *Modern hierarchical, agglomerative clustering algorithms*. 1973, 1–29. <http://arxiv.org/abs/1109.2378>
- O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J. F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., & Sanderson, B. M. (2016). The Scenario

Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*.
<https://doi.org/10.5194/gmd-9-3461-2016>

- Righi, M., Andela, B., Eyring, V., Lauer, A., Predoi, V., Schlund, M., Vegas-Regidor, J., Bock, L., Brötz, B., De Mora, L., Diblen, F., Dreyer, L., Drost, N., Earnshaw, P., Hassler, B., Koldunov, N., Little, B., Loosveldt Tomas, S., & Zimmermann, K. (2020). Earth System Model Evaluation Tool (ESMValTool) v2.0-technical overview. *Geoscientific Model Development*. <https://doi.org/10.5194/gmd-13-1179-2020>
- Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., & Mosher, S. (2013). Berkeley Earth Temperature Averaging Process. *Geoinformatics & Geostatistics: An Overview*. <https://doi.org/10.4172/2327-4581.1000103>
- Sanderson, B. M., Knutti, R., & Caldwell, P. (2015). A representative democracy to reduce interdependency in a multimodel ensemble. *Journal of Climate*, 28(13), 5171–5194. <https://doi.org/10.1175/JCLI-D-14-00362.1>
- Sanderson, B. M., Wehner, M., & Knutti, R. (2017). Skill and independence weighting for multi-model assessments. *Geoscientific Model Development*, 10(6), 2379–2395. <https://doi.org/10.5194/gmd-10-2379-2017>
- Schwalm, C. R., Glendon, S., & Duffy, P. B. (2020). RCP8.5 tracks cumulative CO2 emissions. *Proceedings of the National Academy of Sciences of the United States of America*, 117(33), 19656–19657. <https://doi.org/10.1073/pnas.2007117117>
- Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W., & Bronaugh, D. (2013). Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *Journal of Geophysical Research Atmospheres*, 118(4), 1716–1733. <https://doi.org/10.1002/jgrd.50203>
- Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. In *Bulletin of the American Meteorological Society*. <https://doi.org/10.1175/BAMS-D-11-00094.1>
- Tebaldi, C., Debeire, K., Eyring, V., Fischer, E., & Fyfe, J. (2020). *Climate model projections from the Scenario Model Intercomparison Project (ScenarioMIP) of CMIP6*.
- Tselioudis, G., Lipat, B. R., Konsta, D., Grise, K. M., & Polvani, L. M. (2016). Midlatitude cloud shifts, their primary link to the Hadley cell, and their diverse radiative effects. *Geophysical Research Letters*, 43(9), 4594–4601. <https://doi.org/10.1002/2016GL068242>
- Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., Klein, S. A., & Taylor, K. E. (2020). Causes of Higher Climate Sensitivity in CMIP6 Models. *Geophysical Research Letters*, 47(1), 1–12. <https://doi.org/10.1029/2019GL085782>

Appendix

A2 - Appendix to Science part 2 - CMIP6 sub-selection based on an assessment of key performance criteria (MO)

Table A2. Current available flags for assessed CMIP6 models. Flags are based on classifications in section 3.3 of the main text. Note that the table spans multiple pages.

	All regions								NEU				CEU				MED					
	Seasonal mean blocking frequency (DJF)	Mean DJF circulation	Seasonal mean blocking JJA	Mean JJA circulation	Atlantic SST DJF	Atlantic SST JJA	Storm Track DJf	Storm Track JJA	AMO C	Mean temp bias JJA	Precip annual cycle	Storm Track DJF	Storm Track JJA	Mean temp bias	Precip annual cycle	Storm track DJF	Storm Track JJA	Mean temp bias JJA	Precip annual cycle	Storm Track DJF	Storm Track JJA	
AWI-CM-1-1-MR																						
AWI-CM-1-1-LR																						
ACCESS-CM2																						
ACCESS-ESM1-5																						
BCC-CM2-MR																						
BCC-ESM1																						
CAMS-CSM1-0																						
CESM2-WACCM-FV2																						

Model	All regions										NEU				CEU				MED			
	Seasonal mean blocking frequency (DJF)	Mean DJF circulation	Seasonal mean blocking JJA	Mean JJA circulation	Atlantic SST DJF	Atlantic SST JJA	Storm Track DJf	Storm Track JJA	A M O C	Mean temp bias JJA	Precip annual cycle	Storm Track DJF	Storm Track JJA	Mean temp bias	Precip annual cycle	Storm track DJF	Storm Track JJA	Mean temp bias JJA	Precip annual cycle	Storm Track DJF	Storm Track JJA	
CESM2-WACCM	Yellow	White	Orange	Yellow	Yellow	Yellow	Grey	Grey	Yellow	White	Yellow	Grey	Grey	White	Orange	Grey	Grey	Yellow	Yellow	Grey	Grey	
CESM2	Orange	Yellow	Orange	Yellow	Orange	Yellow	Grey	Grey	White	Yellow	Grey	Grey	Yellow	Yellow	Grey	Grey	White	Yellow	Grey	Grey		
CESM2 – FV2	Red	Red	Orange	Red	Orange	Yellow	Grey	Grey	Yellow	White	Yellow	Grey	Grey	Orange	Grey	Grey	Yellow	Yellow	Grey	Grey		
CMCC-CM2-SR5	Grey	White	Grey	Yellow	Yellow	Orange	Orange	White	Grey	Yellow	Orange	Yellow	Yellow	Orange	Yellow	Yellow	Yellow	Orange	Yellow	Orange	Yellow	
CNRM-CM6-1-HR	White	Yellow	Yellow	Yellow	Yellow	Orange	Yellow	Yellow	White	Yellow	Yellow	Yellow	Yellow	White	Yellow	Orange	Yellow	Orange	Yellow	Orange	Yellow	
CNRM-CM6-1	White	White	Yellow	Yellow	Yellow	Orange	Grey	Grey	Yellow	Yellow	Grey	Grey	Grey	Yellow	Yellow	Grey	Grey	Yellow	Yellow	Grey	Grey	
CIESM	Grey	Yellow	Grey	Yellow	Grey	Grey	Grey	Grey	Yellow	Grey	Grey	Grey	Grey	Orange	Grey	Grey	Grey	Orange	Grey	Grey	Grey	
CNRM-ESM2-1	White	Yellow	Yellow	Yellow	Yellow	Orange	Grey	Grey	Orange	Yellow	Grey	Grey	Grey	Orange	Yellow	Grey	Grey	Yellow	Yellow	Grey	Grey	
CanESM5	Red	Red	White	Orange	Orange	Red	Grey	Grey	Orange	Yellow	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Yellow	Grey	Grey	Grey	
EC-Earth3-Veg	Grey	Yellow	Grey	Yellow	Orange	Grey	White	Yellow	White	Grey	Grey	Yellow	Yellow	Grey	Grey	Yellow	Yellow	Grey	Grey	Yellow	Yellow	
EC-Earth3	White	Yellow	White	Yellow	White	Yellow	White	White	Grey	Yellow	Grey	Yellow	Yellow	White	Yellow	Yellow	Yellow	White	Yellow	Yellow	Yellow	
FGOALS-g3	Red	Red	Yellow	Yellow	Red	Red	Grey	Grey	Red	Orange	Grey	Grey	Grey	Red	Red	Grey	Grey	Orange	Yellow	Grey	Grey	
FIO-ESM-2-0	Grey	Yellow	Grey	Yellow	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Grey	
GFDL-CM4	Grey	Yellow	Grey	Yellow	Yellow	Yellow	Grey	Grey	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Orange	Yellow	Yellow	Yellow	Yellow	Yellow	
GFDL-ESM4	Grey	Yellow	Grey	White	Yellow	Yellow	Grey	Grey	Yellow	Yellow	Grey	Grey	Grey	Yellow	Yellow	Grey	Grey	Yellow	Yellow	Grey	Grey	
HadGEM3-GC31-LL	White	White	Orange	Yellow	Yellow	Yellow	Yellow	Yellow	White	Yellow	Yellow	Yellow	Yellow	White	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Orange	
INM-CM4-8	Yellow	Orange	Red	Orange	Orange	Orange	Grey	Grey	Orange	Yellow	Grey	Grey	Grey	Red	Red	Grey	Grey	Red	Yellow	Grey	Grey	

GISS-E2-1-G	Red	Orange	White	Grey	Yellow	Orange	Red	Yellow	Grey	Orange	Orange	Orange	Yellow	Orange	Orange	Yellow	Yellow	Orange	Yellow	Orange	Yellow
HadGEM3-GC31-MM	White	White	White	Yellow	Yellow	White	White	White	Grey	Grey	Yellow	Yellow	Grey	Grey	Yellow	Yellow	Grey	Grey	White	Yellow	
UKESM-0-LL	White	White	Yellow	Grey	Yellow	White	White	Orange	White	Yellow	Grey	Grey	White	Yellow	Grey	Grey	Yellow	Yellow	Grey	Grey	
EC-Earth3-AerChem	Grey	Grey	Grey	Grey	Orange	Yellow	White	White	White	Yellow	Yellow	Yellow	White	Yellow	Yellow	Yellow	White	Yellow	Yellow	White	
EC-Earth3-CC	Grey	Grey	Grey	Grey	Grey	Yellow	White	White	Grey	Grey	Yellow	Yellow	Grey	Grey	Yellow	Yellow	Grey	Grey	Yellow	Yellow	
KOST-ESM	Grey	Grey	Grey	Grey	Grey	Grey	Orange	White	Grey	Grey	Yellow	Orange	Grey	Grey	Yellow	Yellow	Grey	Grey	Orange	Yellow	
KACE-1-0-G	White	Grey	Yellow	Grey	Yellow	Yellow	White	White	Orange	Grey	Yellow	Yellow	Red	Grey	Yellow	Orange	Orange	Grey	Orange	Yellow	
E3SM-1-1-ECA	Grey	Grey	Grey	Grey	Red	Red	Grey	Grey	Orange	Grey	Grey	Grey	Orange	Grey	Grey	Grey	Orange	Grey	Grey	Grey	
NorCPM1	Grey	Grey	Grey	Grey	Orange	Orange	Grey	Grey	Orange	Grey	Grey	Grey	Orange	Grey	Grey	Grey	Orange	Grey	Grey	Grey	
Fgoals-f3-L	Grey	Grey	Grey	Grey	Red	Red	Red	Orange	Orange	Yellow	Grey	Grey	Grey	Yellow	Grey	Grey	Orange	Grey	Grey	White	

Appendix

A3 - Appendix to Science part 4 - Selecting regional simulations tailored to user's needs and setting small (CPRCMs) and large (GCMs) ensembles into uncertainty context (SMHI)

Table A3a. GCM simulations (rcp85) used in the second study, downloaded from ESGF.

GCM simulations	runs rip	GCM simulations	runs rip
ACCESS1-0	r1i1p1	GISS-E2-R-CC	r1i1p1
ACCESS1-3	r1i1p1	HadGEM2-AO	r1i1p1
bcc-csm1-1	r1i1p1	HadGEM2-CC	r1i1p1, r2i1p1, r3i1p1
bcc-csm1-1-m	r1i1p1	HadGEM2-ES	r1i1p1, r2i1p1, r3i1p1, r4i1p1
BNU-ESM	r1i1p1	inmcm4	r1i1p1
CanESM2	r1i1p1, r2i1p1, r3i1p1, r4i1p1, r5i1p1	IPSL-CM5A-LR	r1i1p1, r2i1p1, r3i1p1, r4i1p1
CCSM4	r1i1p1, r2i1p1, r3i1p1, r4i1p1, r5i1p1, r6i1p1,	IPSL-CM5A-MR	r1i1p1
CMCC-CESM	r1i1p1	IPSL-CM5B-LR	r1i1p1
CMCC-CM	r1i1p1	MIROC5	r1i1p1, r2i1p1, r3i1p1
CMCC-CMS	r1i1p1	MIROC-ESM-CHEM	r1i1p1
CNRM-CM5	r1i1p1, r2i1p1, r4i1p1, r6i1p1	MIROC-ESM	r1i1p1
CSIRO-Mk3-6-0	r1i1p1, r2i1p1, r3i1p1, r4i1p1, r5i1p1, r6i1p1, r7i1p1, r8i1p1, r9i1p1	MPI-ESM-LR	r1i1p1, r2i1p1, r3i1p1
EC-EARTH	r1i1p1, r2i1p1, r8i1p1, r9i1p1, r12i1p1	MPI-ESM-MR	r1i1p1
FGOALS-g2	r1i1p1	MRI-CGCM3	r1i1p1
FIO-ESM	r1i1p1, r2i1p1, r3i1p1	MRI-ESM1	r1i1p1
GFDL-CM3	r1i1p1	NorESM1-ME	r1i1p1
GFDL-ESM2G	r1i1p1	NorESM1-M	r1i1p1
GFDL-ESM2M	r1i1p1	KNMI-EC-EARTH	r14i1p1
GISS-E2-H-CC	r1i1p1		
GISS-E2-H	r1i1p1, r1i1p2, r1i1p3, r2i1p1, r2i1p3		

Table A3b. List of RCMs and their downscaled GCMs.

GCM \ RCM	CNRM-CERFACS-CNRM-CM5	ICHEC-EC-EARTH			IPSL-IPSL-CM5A-MR	MOHC-HadGEM2-ES	MPI-M-MPI-ESM-LR			NCC-NorESM1-M	CCCma-CanESM2
	(r1)	(r1)	(r3)	(r12)	(r1)	(r1)	(r1)	(r2)	(r3)	(r1)	(r1)
CLMcom-CCLM4-8-17_v1	85	85	85	85	85	85	85	85	85	85	85
CLMcom-ETH-COSMO-crCLIM-v1-1-v1	1	1	1	1		1	1	1	1	1	1
CNRM-ALADIN63_v1 / _v2	1					1	1	1		1	
DMI-HIRHAM5_v1 / _v2 / _v3	1	1	1	1	1	1	1			1	
MPI-CSC-REMO2009_v1							1	1			
GERICS-REMO2015_v1 / _v2	1			1	1	1			1	1	1
ICTP-RegCM4-6_v1	1			1		1	1			1	
KNMI-RACMO22E_v1 / _v2	1	1	1	1	1	1	1	1		1	
MOHC-HadREM3-GA7-05_v1				1		1	1			1	
SMHI-RCA4_v1 / _v1a		1	1	1	1	1	1	1	1	1	
IPSL-WRF381P_v1 / _v2	1			1	1	1	1	1		1	

Table A3c. List of CPRCMs with their respectively downscaled RCMs and GCMs.

CPRCM name	downscaled RCM	downscaled GCM
CLMcom-CMCC-CCLM5-0-9	CLMcom-CMCC-CCLM5-0-9	EC-EARTH
CNRM-AROME41t1	CNRM-ALADIN63	CNRM-CERFACS-CNRM-CM5
ETHZ2-COSMO-pompa_5.0	COSMO-pompa_5.0	MPI-M-MPI-ESM-LR
GERICS-REMO2015	GERICS-REMO2015	MPI-M-MPI-ESM-LR
HCLIMcom-HCLIM38-AROME	HCLIM38-ALADIN	EC-EARTH (r12)
ICTP-RegCM4-7	RegCM4-7	mohc-hadgem2-es
IPSL-WRF381CE	WRF381CE	IPSL-CM5A-MR
KNMI-HCLIM38h1-AROME	KNMI-RACMO23E	KNMI-EC-EARTH (r14, r13)

Appendix

A4 - Appendix to Science part 6 - Pseudo-global warming experiments to create event sets of future weather (KNMI)

The below is a paper in preparation for submitting to a scientific journal. Note that the numbering of figures and tables within this paper (i.e. Appendix A4) is independent from the rest of the deliverable, Figure 1 in the text here refers to Figure 1 of this appendix.

The 2018 west-central European drought projected in a warmer climate

E.E. Aalbers, E. van Meijgaard, G. Lenderink, H. de Vries, B. van den Hurk

A4.0 ABSTRACT

Projections of changes in extreme droughts are associated with large uncertainties, owing to the complex genesis of droughts, uncertainty in the atmospheric dynamics, such as the representation of atmospheric blocking conditions and land-atmosphere feedbacks in climate models. In this study we investigate the impact of global warming on soil moisture drought severity in west-central Europe by employing so-called pseudo- global warming (PGW) experiments, which set the 1980-2020 period against the back- ground of a globally warmer world. We thus target changes in drought severity conditional on the day-to-day evolution of the atmospheric circulation in the historical period, by providing future analogues of present-day drought episodes.

The 2018 west-central European drought is the most severe drought in the 1980-2020 period in this region. Under 1.5K, 2K and 3K global warming, the 2018 drought episode experiences strongly enhanced summer temperatures, but a fairly modest response in soil moisture availability and small changes in drought duration. This is owing to strongly constrained evaporation in summer under present-day and future conditions, and increasing precipitation in spring, autumn and winter. For 2018 conditions, the drought severity, defined as the cumulative soil moisture deficit volume during a drought episode, increases with 17% to 34% under 1.5K to 3K global warming. In the 1980-2020 period under +2K warming however, increasing evaporation and decreasing summer precipitation result in a strong increase in drought frequency and severity, with 2019/20-like drought conditions occurring once every three years. Strikingly, many years without noticeable droughts in the historical period show very strong drying responses, and exceed 2003-like drought conditions under +2K global warming, accompanied by regional temperatures considerably exceeding +2K.

Our results thus provide examples of future droughts, illustrate the strongly enhanced drought risk in west-central Europe under global warming, even without taking into account the full range of dynamical responses to climate change, and may contribute to the formulation of effective adaptation strategies.

A4.1 INTRODUCTION

The impact of recent west-central European heat waves on society and nature (Vogel et al., 2019, Rösner et al., 2019, Kramer et al., 2019, Krikken et al., 2019, Schuldt et al., 2020, Beillouin et al., 2020) once again triggered questions regarding the role of climate change in the occurrence and extremity of drought events (Kornhuber et al., 2019, Yiou et al., 2020, Philip et al., 2020, Zscheischler and Fischer,

2020) and on what to expect under continuing global warming (Toreti et al., 2019, Kornhuber et al., 2019, Hari et al., 2020).

The 2018 growing season was the compound hottest-and-driest ever recorded in west-central Europe (Toreti et al., 2019, Zscheischler and Fischer, 2020), owing to a sequence of anomalously persistent high pressure systems over eastern, northern, and central Europe between April and October (Bissolli, 2019, Sluijter et al., 2018), associated with subsidence, clear sky conditions and generally low relative humidity and moisture advection (Sousa et al., 2017, 2018), against the background of globally increasing temperatures (Philip et al., 2020, Vogel et al., 2019). Temperatures were anomalously high over almost the entire European continent (Vogel et al., 2019, Kornhuber et al., 2019), but the precipitation deficit was particularly intense and long lasting in west-central Europe, where it built up from April/May until November, only intermittently interrupted by intense but small-scale short-duration rainfall events (Bissolli, 2019, Sluijter et al., 2018). This led to soil desiccation and extremely low levels of groundwater tables (Brakkee et al., 2021) and river discharge in the west-central European river basins (Brunner et al., 2019, Kramer et al., 2019).

Although the probability of heat waves in this region is demonstrated to have increased in response to anthropogenic climate change (Stott et al., 2004, Vogel et al., 2019, Vautard et al., 2020), the attribution of extreme drought conditions is complicated by the complexity of processes contributing to wide-spread drought conditions (Shepherd, 2014, Trenberth et al., 2014), while statistics are complicated by the scarcity of independent drought events owing to their long timescale and large spatial scale. Intense drought conditions are governed by persistent patterns of atmospheric circulation with low moisture advection into the region of interest. Trends over recent years suggest increases in the frequency and/or persistence of such circulation conditions (Coumou et al., 2014, Kornhuber et al., 2019), but there are no significant long-term trends in meteorological (precipitation) drought events in this region (Gudmundsson and Seneviratne, 2016, Hanel et al., 2018, Manning et al., 2019, Philip et al., 2020). Nevertheless, observation- and model-based studies find decreasing trends in summer water availability (precipitation minus evaporation) which are attributable to global warming (Padrón et al., 2020), increases in the frequency and/or severity of soil moisture droughts (Hanel et al., 2018, Philip et al., 2020) and in long-duration compound hot-and-dry events (Manning et al., 2019). This is owing to increasing trends in atmospheric evaporative demand with global warming in the predominantly energy-constrained evaporation regime in west-central Europa.

Under further increasing greenhouse-gas concentrations, climate projections agree on a general pattern of year-round decreasing precipitation in the Mediterranean and increasing precipitation in northern Europe, with the drying/wetting transition zone shifting north in summer under higher levels of global warming (Jacob et al., 2014, Coppola et al., 2021, Aalbers et al., 2018). For west-central Europe small decreases or increases are projected for spring and summer, but results are uncertain and depend on e.g. the RCM/GCM resolution and generation (Jacob et al., 2014, Coppola et al., 2021, Van der Linden et al., 2019), biases in the mean climate state in the reference period (Orth et al., 2016, Van der Linden et al., 2019) and the ability of climate models to realistically represent land-surface-atmosphere coupling (Vogel et al., 2018, Selten et al., 2020) and atmospheric dynamics (Shepherd, 2014, Woollings et al., 2018). Studies using large climate model ensembles find further decreases in soil moisture, with strongest responses in summer and autumn (Ruosteenoja et al., 2018, Van der Linden et al., 2019) and increases in the frequency and severity of (multi-year) drought episodes (Samaniego et al., 2018, Toreti et al., 2019, Hari et al., 2020).

In this study the contribution of global warming to the increase in drought severity is being addressed by putting the 1980-2020 historical period in the context of a globally warmer world. We hereby follow a pseudo global warming (PGW) approach ((Schär et al., 1996, Attema et al., 2014, Prein et al., 2017, Kröner et al., 2017, Lenderink et al., 2019)), by which we create “future weather analogues” (Hazeleger et al., 2015, Shepherd, 2019, Sillmann et al., 2019, Van der Wiel et al., 2021, Wehrli et al., 2020) of present-day summers. We focus on the 2018 drought episode for its recent occurrence and severe impact. Based on regional climate model (RCM) simulations driven by reanalysis data, we first explore the atmospheric drivers and soil moisture evolution of the 2018 event under present-day conditions. We repeat this analysis after perturbing the same synoptic conditions in order to match the climate change signal derived from three different global climate model (GCM) projections and for several warming levels. In addition to diagnosing the response in atmospheric drivers, the soil moisture evolution and the severity of the 2018 drought event, we also evaluate the position of this 2018 event in the 1980-2020 period, both for present-day and for future conditions of a single warming level.

The purpose of this work is to provide robust, physically consistent scenarios of what global warming entails for both extreme droughts, and for the full range of wet to moderately dry years that occurred in the historical record. It is intended to complement projections of changes in drought risk derived with the conventional large-ensembles approaches, giving an explicit reference to collectively experienced real world events.

A4.2 MODEL AND METHODS

A4.2a -- Regional climate model

All simulations are performed with the RCM KNMI-RACMO2 (Van Meijgaard et al., 2012), run at 12km resolution, with 40 vertical model levels. External forcings for aerosols and greenhouse gases have been implemented according to CMIP5 prescriptions (Collins et al., 2013). RACMO2 uses the land surface scheme HTESSEL (Balsamo et al., 2009), which employs four soil layers with a total depth of 2.9 m. At the bottom of the soil column, boundary conditions are specified as zero-heat flux and free drainage. Each land-grid cell includes separate tiles for high and low vegetation (16 vegetation types), bare soil, snow on low vegetation/bare soil, snow beneath high vegetation and intercepted water, for which the energy and water balances are solved individually. The tile fractions vary with interception storage, snow cover and a fixed seasonally changing vegetation cover (leaf area index climatology). The model domain is centered over west-central Europe, and covers the majority of Europe.

A4.2b -- Experimental setup

The analyses are based on two sets of RCM simulations: present-day simulations (REF) and pseudo-global-warming (PGW) simulations. Both sets include a climate run covering the period 1980-2017 for present-day conditions and 2K global warming (climREF and clim+2K) and an 11-member initial-condition ensemble for the period 2018-2020 for present-day conditions and 1.5K, 2K and 3K global warming (2018REF and 2018+nK; n=1.5,2,3). By creating an ensemble, random small-scale variations in the weather (i.e. due to internal variability within the RCM domain) are sampled, yielding a more robust assessment of future changes. This is especially relevant for the analysis of single weather or climate events. Tab. 1 provides an overview of all simulations. The simulations are detailed in the following subsections.

Table 1. Model simulations

Present-day simulations (REF)						PGW simulations ($+nK$)				
						Δ_{nK} : perturbation for nK global warming, $n=1.5, 2, 3$				
						EC: EC-EARTH Had: HadGEM2-ES MPI: MPI-ESM-LR				
	Member	Start date	Period	Name	Land surface init.	Sea & atm. init. & forcing	Name	Land surface init.	Sea & atm. init. & forcing	
clim	-	1 Jan	1979-2017	climREF	ERA5	ERA5	clim+2K EC/Had/MPI	ERA5 + Δ_{2K}	ERA5 + Δ_{2K}	
	1	1 Jan	2018-2020	2018REF-mb1	climREF	ERA5	2018+ nK EC/Had/MPI $n=1.5, 2, 3$	mb1	2018REF-mb1 + Δ_{nK}	
2018	2	6 Jan		2018REF-mb1				2018+ nK -mb1		ERA5 + Δ_{nK}
	⋮	⋮		⋮	⋮			⋮	⋮	⋮
	11	20 Feb		2018REF-mb11	2018REF-mb1			mb11	2018+ nK -mb1	

Present-day simulations First, RACMO is run continuously over the period January 1st 1979 – January 1st 2018, with initial conditions and lateral and sea surface boundary conditions from the ERA5 reanalysis dataset (Hersbach et al., 2020). The sea surface and lateral boundary conditions are updated every 3 hours. The first year is used as spin-up, leaving the period 1980-2017 as the reference period for the present-day climate (climREF). The 11-member ensemble for 2018-2020 (2018REF) is created by running RACMO 11-times over the period 2018-2020, reinitializing the atmospheric state to the ERA5 reanalysis at January 1st for member 1, January 6th for member 2, up to February 20th for member 11. Unless indicated otherwise, throughout this paper analyses are based on the ensemble mean values of the 2018REF ensemble. The simulated 2-m temperature and precipitation are evaluated against the gridded observational dataset E-OBS v20.0e (Cornes et al., 2018).

PGW-simulations To examine the impact of global warming, all simulations are rerun, but with perturbed initial (soil moisture, soil temperature, snow cover), sea surface (temperature and sea ice extent) and lateral boundary conditions (temperature, humidity, wind), representing the change in the mean climate state in a globally warmer world. We impose a single 2K global warming to the 1979-2018 simulation (clim+2K), and three different global warming levels to the 11-member 2018 ensemble (2018+1.5K, 2018+2K and 2018+3K) to examine the sensitivity of the 2018 response to the warming level. The perturbations are determined from global climate model (GCM) projections as the 3-dimensional monthly mean climate change signal in the 30-year period in which the target global warming level is reached with respect to present-day conditions (1991-2020), see Tab. 1. These include thermodynamic changes, changes in the vertical temperature and humidity profile, and monthly mean changes in the circulation. Therewith we capture a large part of the climate change signal (Brogli et al., 2019,b), but the day-to-day evolution of the synoptic-scale circulation, i.e. the sequence of weather systems entering the model domain in the PGW simulations remains essentially determined by the reanalysis forcing and is therefore very similar to the sequence seen in the present-day (or control) simulation (Brogli et al., 2019,b, Prein et al., 2017).

Since climate models differ in their regional climate response given a certain level of global warming, we derive three sets of perturbations from three different GCM initial- condition ensembles: a 16-member EC-EARTH v2.3 (Hazeleger et al., 2010) ensemble produced at KNMI, a 4-member HadGEM2-ES (Collins et al., 2011) ensemble and a 3-member MPI-ESM-LR (Giorgetta et al., 2013) ensemble from the CMIP5 archive (Taylor et al., 2012), referred to as respectively EC, HAD and MPI. The perturbations are derived from initial-condition GCM ensembles rather than from a single simulation per GCM to obtain a robust estimate of the forced climate response (Deser et al., 2010, Fischer et al., 2014, Aalbers

et al., 2018). All GCM ensembles are run under the high-end emission scenario (RCP8.5). HAD exhibits the strongest warming in spring, MPI shows the strongest warming and drying in summer.

A4.2c -- Indicators and variables

We identify soil moisture drought conditions based on the exceedance of a seasonally varying threshold of the soil wetness index (SWI, (-)) of the top 1 m of the soil. The top 1 m of the soil is where – in HTESSEL – vegetation has the highest root density and where water deficiencies have the strongest link to agricultural drought (Seneviratne et al., 2012). The SWI is the fraction of plant available water in the top 1 m of the soil, defined as the soil moisture availability (θ , (mm)) scaled between field capacity (θ_{fc}) and permanent wilting point (θ_{pwp}) (Eq. 1). The SWI is better suited for aggregation over areas with different soil types than θ itself. θ_{fc} and θ_{pwp} are fixed characteristics per grid cell.

$$SWI = \frac{\theta - \theta_{pwp}}{\theta_{fc} - \theta_{pwp}} \quad (1)$$

A soil moisture drought event is defined as the consecutive period in which the soil moisture conditions are drier than the 5th percentile threshold of the 1980-2017 SWI (SWI5th). SWI5th is calculated for every calendar day based on 14-day smoothed SWI values (Trenberth et al., 2014). We apply the same drought threshold for the present-day and PGW conditions, to benchmark the warming induced changes to present-day conditions. We express the drought severity in terms of the drought deficit volume (unit mm d), which integrates drought duration (unit d) and drought intensity (unit mm). It is calculated as the accumulated difference between θ_{5th} and θ over the drought episode. The drought intensity is defined as the drought deficit volume divided by the drought duration.

In the analyses we use the atmospheric evaporative demand synonymously with potential evaporation (E_p), the evaporation that would take place assuming unconstrained conditions with respect to soil moisture availability and vapor pressure deficit. The computation of evaporation in HTESSEL uses a resistance approach based on Jarvis (1976), for each individual land cover tile, see ECMWF (2009). To obtain a potential evaporation measure that is fully consistent with the simulated actual evaporation, it is diagnosed in a parallel calculation within RACMO2, using the prevailing atmospheric conditions, but with resistance functions accounting for soil moisture availability in the top layer (soil evaporation), soil moisture availability in the root zone (transpiration) and vapor pressure deficit (transpiration) set at 1, representing unconstrained conditions. See for details supplementary material B.

A4.2d -- Study area

To study the local intensification of the 2018-drought episode and accompanying heat wave we have selected the larger river basins in west-central Europe that discharge in the North Sea, namely the Rhine, Meuse, Scheldt, Ems, Weser and Elbe. These river basins are part of the area where the 2018 soil moisture drought episode was most severe and lasted longest, as shown in Fig. 1.

A4.3 -- THE 2018 DROUGHT EPISODE IN THE PRESENT-DAY CLIMATE

We first present the main characteristics of the simulated 2018 drought episode, and briefly discuss the evaluation of the temperature and precipitation anomalies against observations. In Fig. 1a-d maps of the simulated 2018 anomaly in 500 hPa geopotential height (contours), temperature, precipitation, evaporation and soil moisture are shown for April-June, July-September and October - December.

Anomalies are calculated from 2018REF with respect to the 1980-2017 period (climREF). Time series of these variables averaged over the west-central European river basins are shown in right panel, Fig. 1e-h, with observed temperature and precipitation delineated in green.

The high pressure anomalies in (late) spring (AMJ) and summer (JAS), clearly co- occur with the large positive temperature anomalies and high precipitation deficits. Averaged over the west-central European river basins the simulated (observed) temperature anomaly is +3.1°C (+2.5°C) over the growing season (April to September), and temperatures exceed the 95th percentile during several episodes (Fig. 1e). Most noteworthy are 8 - 22 April, with a 15-day mean temperature anomaly of +6.5 °C (+6.0 °C), and 22 July – 8 August, with an 18-day mean anomaly of +5.9 °C (+5.1 °C). The latter period was indeed classified as heat wave in the individual countries (Yiou et al. 2019, Sluijter et al. 2018, Vogel et al. 2019, Bissolli et al. 2019). Apart from a cold bias in winter, the basin-mean simulated absolute temperatures are fairly accurate, with a small underestimation with respect to the observed 1980-2017 mean temperature in the growing season (-0.3 °C) and overestimation of the extreme conditions of 2018 temperatures (+0.3 °C) in most members of the 2018REF ensemble.

Basin-mean precipitation is anomalously low in each month from February to November, while locally anomalously high precipitation occurs in the coastal region in April, and in the western part of the basin in August. Averaged over the growing season, the simulated (observed) basin-mean precipitation anomaly is -1.1 mm/d or -41% (-37%), with largest, basin-wide deficits in June (-56% (-43%)) and in July (-64% (-64%)). Precipitation is slightly overestimated compared to the observations with on average 0.2 mm/day, both for the climatology and 2018. The higher precipitation amounts could imply an overestimation of the soil moisture levels. However, the overestimation of the 2018 summer temperature suggests slightly too dry conditions. Apart from model biases, this is possibly related to water management and interactions between soil moisture and groundwater in especially the low-lying coastal areas, which are not taken into account in HTESSSEL.

Under prevailing conditions of clear skies, high solar radiation, high temperatures and increasingly dry air, the atmospheric evaporative demand is extremely high (+1 mm/d or +35% over the growing season, see Fig. 1g). Also the actual evaporation is anomalously high from April up until beginning June, modulating the near-surface temperatures (Fig. 1b,g). However, it cannot keep up with the rise in atmospheric evaporative demand, owing to quickly increasing soil and canopy resistance against evaporation in response to decreasing relative humidity and soil moisture availability (Fig. 1h), and has below normal values from mid-June to October. As a consequence, the sensible heat flux strongly increases (not shown), which corresponds to an amplified rise in summer and autumn near-surface temperatures. Averaged over the growing season, the actual evaporation is slightly smaller than normal (-0.1 mm/day or -6%).

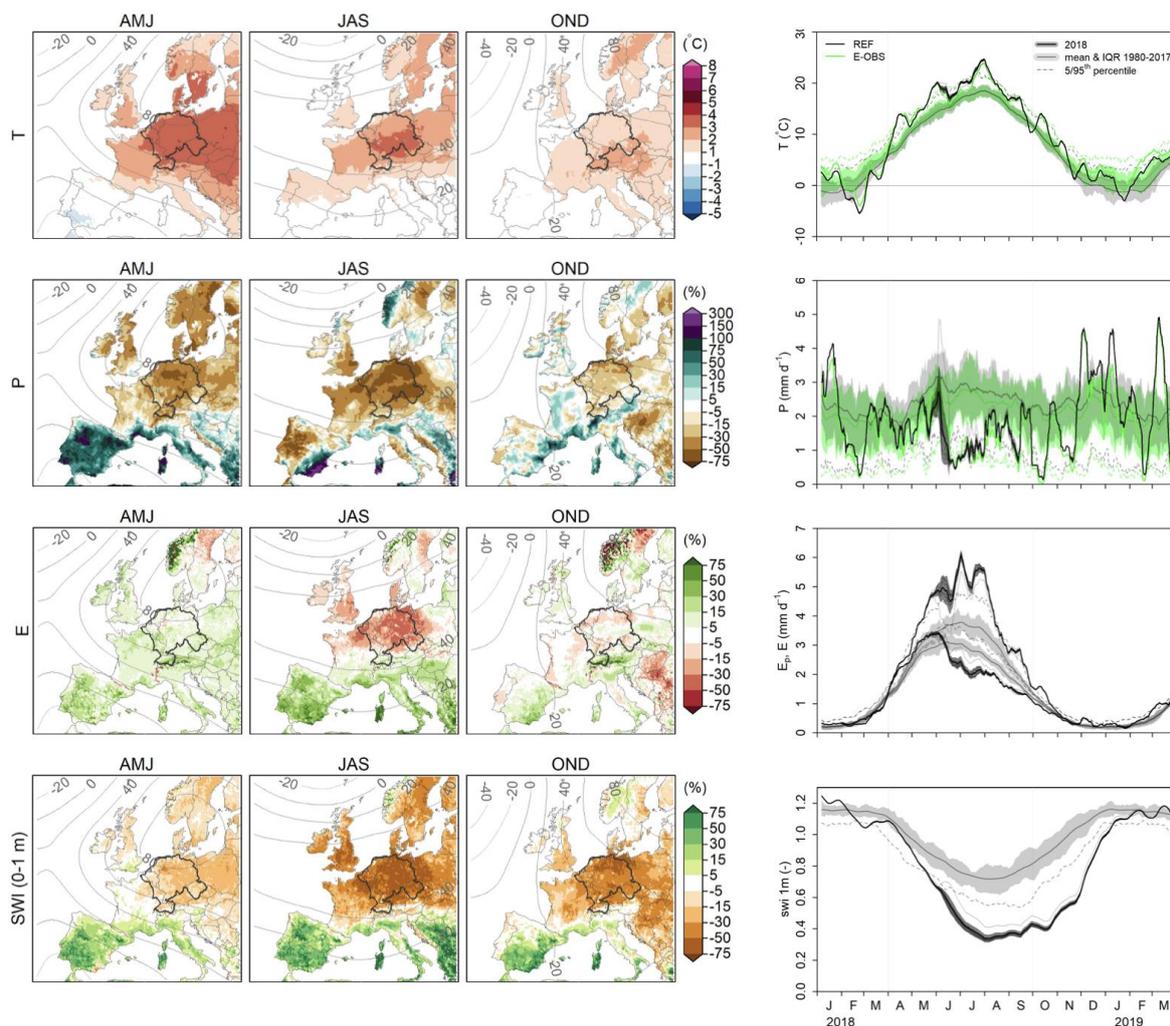


Figure 1: The 2018 drought episode in simulations (2018REF and climREF) and observations (E-OBS v20.0). Maps show the seasonal mean anomaly in 500 hPa geopotential height (contours) and a) near- surface temperature, b) precipitation, c) evaporation and d) SWI (top 1 m) from April to December 2018 with respect to 1980-2017 and the basin-mean time series from January 2018 to March 2019 and climatology for the same variables (e-h). Date is masked over sea for visibility. Potential evaporation is shown along with the actual evaporation (g). Observed temperature (e) and precipitation (f) are shown in green. The dark lines and shading show the 2018REF ensemble mean, the individual members and the ensemble spread (± 1 standard deviation). The lighter line and shading and the dashed line depict respectively the 1980-2017 mean, interquartile range, and the 5th or 95th percentile depending on the variable. Time series are smoothed with a 14-day running mean.

The resulting extremity of the 2018 soil moisture drought is clearly reflected in Fig.1d,h. Anomalously low soil moisture levels occur in large parts of central and northern Europe, but, consistent with the persistent precipitation deficits, conditions are most severe and persistent in west-central Europe. Averaged over the west-central European river basins the soil moisture conditions are around normal at the start of the growing season, owing to low temperatures and evaporation in March, but soils steadily deplete from April onwards, reach severely dry conditions (exceeding the 5th percentile) in the second half of May and minimum soil moisture levels in early August. Soil moisture levels remain very low throughout the growing season up to end October. Only then precipitation starts to exceed the evaporation and soil moisture replenishes, reaching the 5th percentile threshold in the beginning of January 2019, after nearly 8 months of severely dry conditions. By then, the soil moisture deficit

volume over the nearly 8 months is 8240 mm d, with a mean drought intensity of 36 mm. Normal soil moisture levels in the top 1 m of the soil are reached early February 2019. For deeper soil layers the winter precipitation is insufficient to fully replenish the soils to normal levels, and the anomalously dry conditions persist throughout 2019.

The spread in the 2018REF ensemble is largest in the period end May-early June when spread in the location and intensity of precipitation bearing systems can induce relatively strong variability in wetness and temperature. While the ensemble spread in temperature is relatively short-lived, the ensemble spread in soil moisture reduces more gradually over summer. One ensemble member receives relatively high precipitation amounts in May – early June, as well as in July and August (Fig. 1f). Evaporation in this member is consequently relatively high throughout the summer (Fig. 1g), and, hence, temperature is lower than the ensemble mean, ranging from 1.1K in June to 0.3K in September (Fig. 1e).

A4.4 -- THE 2018 DROUGHT EPISODE IN THE PRESENT-DAY CLIMATE

A4.4a -- Climatological mean response to 2K warming

To provide context to the drought years, we first present the climatological mean response to a 2K warming. Fig. 2 shows the seasonal response patterns in geopotential height (contours), near-surface temperature, precipitation, (potential) evaporation, and soil moisture over Europe for the EC-perturbed simulations (see Tab. 1). The annual cycle in the basin-mean response in these and additional variables is shown in Fig. 3.

The spatial response patterns exhibit the well-known seasonally-varying warming and drying gradients over Europe (e.g. Coppola et al., 2021), showing that the PGW- simulations indeed capture the main characteristics of the full climate response (Brogli et al., 2019b). In spring, autumn and winter the warming gradient is oriented roughly northwest-southeast, with weakest warming above the British Isles and coastal regions adjacent to the Atlantic ocean and Northsea (blue colors represent below 2K warming). In summer, warming ranges between around +2.0K over Scandinavia to around +3.0K, locally +3.5K in southern Europe. Averaged over the river basins the near-surface temperature response varies between +1.4K in May and +2.6K in August (Fig. 3a). Note that the spread around the 1980-2017 mean response is rather large, especially in JAS, which will be discussed in Sec. 5.2.

The transition zone of increasing precipitation in the north and decreasing precipitation in the south is positioned just southwest of the west-central European river basins in spring and autumn and over the northeast of the basins in summer, yielding increasing precipitation in winter, autumn and early spring, and small decreases in summer for the basin-mean (Fig. 2b, 3e). This co-occurs with nearly constant relative humidity and increasing cloud cover in late autumn, winter and early spring, decreases in relative humidity and cloud cover in JJASO, and consequent increases in net surface solar radiation in this period (Fig. 3b,d). Under conditions of higher temperatures, and enhanced by the increase in solar radiation and decrease in relative humidity from late spring to late autumn, the atmospheric evaporative demand increases over land throughout the year (Fig. 2c, 3f). Note that decreases in evaporative demand are found over the Atlantic owing to a decrease in the ocean-air humidity gradient (Lainé et al., 2014). The present-day soil moisture regime in west-central Europe allows for year-through increases in actual evaporation in almost all years, with around potential rate

in winter and early spring, but smaller than potential in JJASO (Fig. 2d, 3g) resulting in increases in the sensible heat flux in the latter period (Fig. 3c).

For the combined river-basin area, the year-through enhanced evaporation and reduced summer precipitation lead to enhanced soil moisture depletion in late spring and summer, while in autumn and winter increases in precipitation result in a faster soil moisture replenishment. This is enhanced by reduced snowmelt in spring and a larger fraction of precipitation falling as rain in autumn (not shown). The resulting soil moisture levels in the top 1 m of the soil are around present-day or even wetter conditions in winter and early spring, but drier from mid-June to December, with a maximum drying response in September (Fig. 2e, 3h). In summer and autumn, the soil moisture availability in deeper layers and runoff decrease as well (not shown). However, the response in annual precipitation equals the response in annual evaporation, meaning that each winter soil moisture levels in all layers are restored to present-day levels, and decreases in summer runoff are compensated by increases in winter.

With the amplitude of the response and the position of the drying/wetting transition zone being dependent on the GCM, the MPI- and HAD-perturbed simulations give slightly different results, see Fig. C2 and C1. The drying/wetting transition zone is located further northeast in all seasons for both clim+2K|MPI and clim+2K|HAD. clim+2K|MPI shows a weaker temperature response in spring, but much stronger drying and warming in JAS, consistent with a strong response in the geopotential height anomaly, relative humidity, cloud cover and solar radiation. The soil moisture depletion and decrease in runoff over the growing season are stronger, but so is the soil moisture replenishment, and present-day soil moisture levels are reached in late winter. Annual runoff however decreases in clim+2K|MPI. clim+2K|HAD gives a much weaker increase or even decrease in cloud cover and relative humidity in autumn, winter and spring than the EC- and MPI-perturbed simulations, the spring temperature and solar radiation response are consequently stronger, and so is the increase in spring evaporation. Moreover, increases in precipitation are overall smaller, and soil moisture levels are found to decrease earlier in spring and decreases in runoff occur throughout the year.

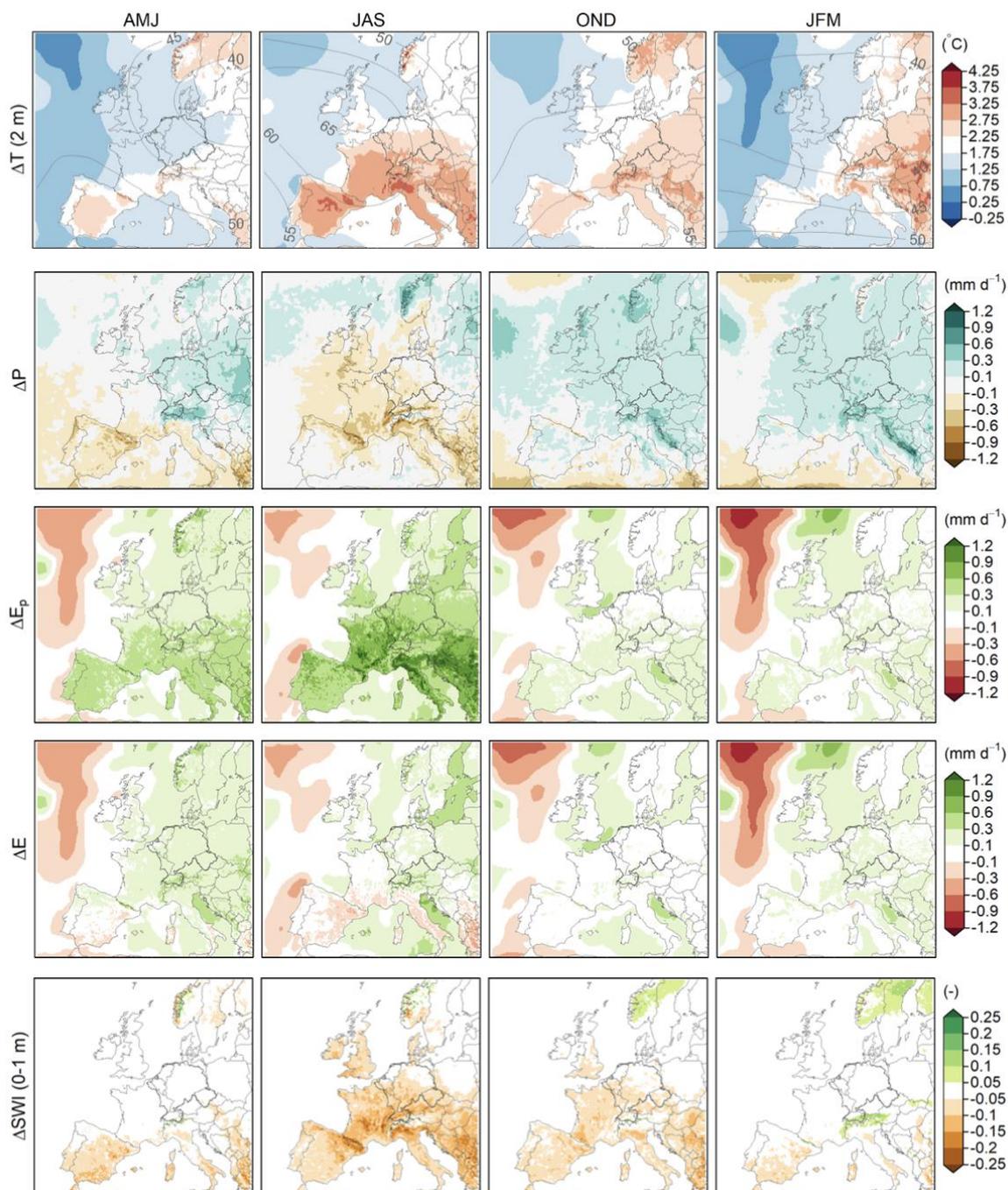


Figure 2: Climatological mean response (1980-2017) to 2K warming in a) the geopotential height at 500hPa (contours) and near-surface temperature (T, shading), b) precipitation (P), c) potential evaporation (Ep), d) evaporation (E) and e) the soil wetness index in the top 1m of the soil, averaged over April - June (AMJ), July - September (JAS), October - December (OND) and January - March (JFM). Results are based on climREF and clim+2K|EC.

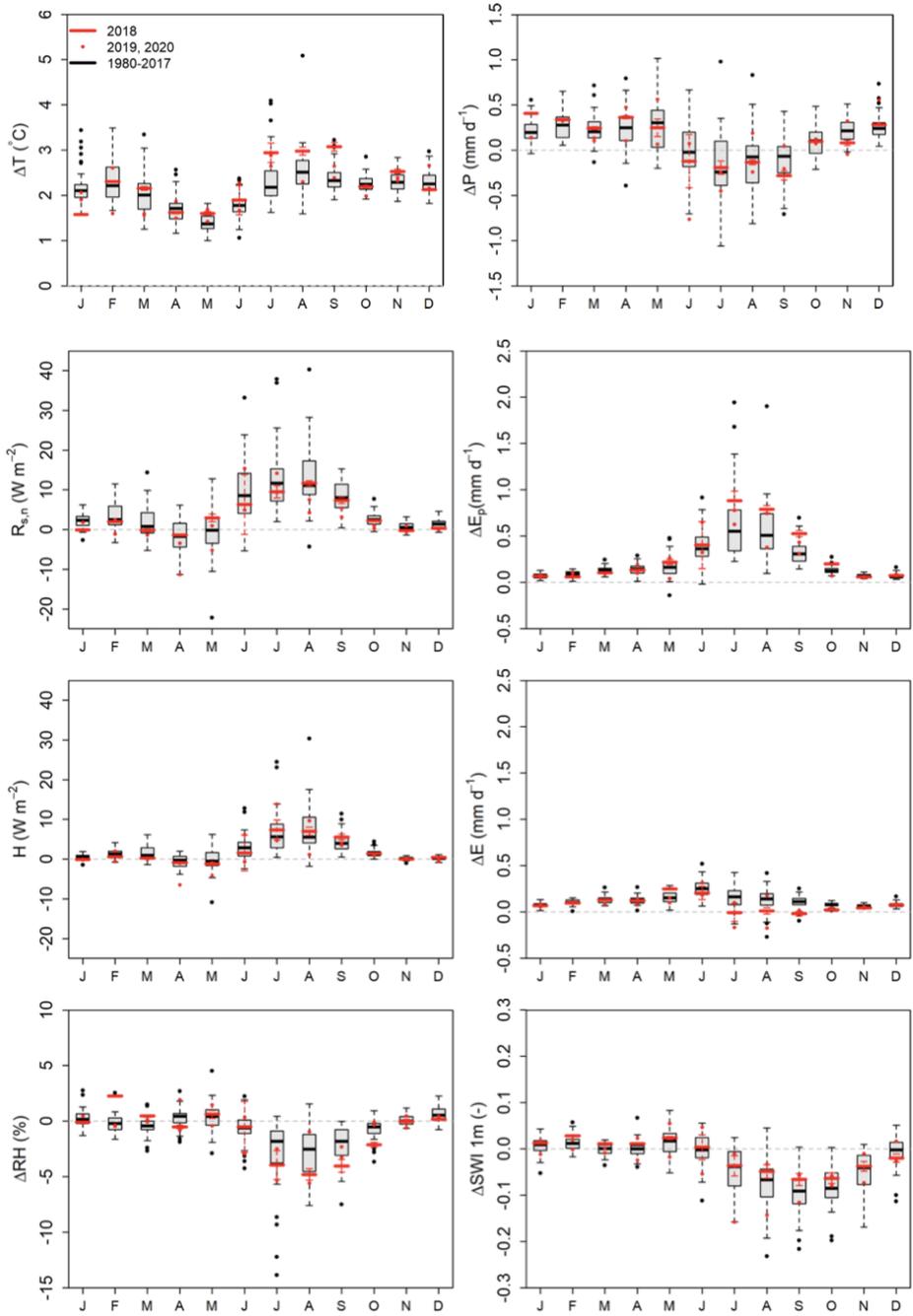


Figure 3: Annual cycle in the basin-mean response to 2K global warming in a) near-surface temperature, b) net solar radiation at the surface ($R_{s,n}$), c) sensible heat flux (H), d) near-surface relative humidity (RH), e) precipitation, f) potential evaporation, g) evaporation and h) SWI in the top 1 m of the soil. The boxplots show the interannual distribution of the 1980-2017 response, depicting the median (black bar), interquartile range (box), the total range with a maximum distance of 1.5 times the interquartile range outside the box (whiskers), and outliers (black dots). The red bars delineate the response for 2018, the red whiskers the inter-member spread in the response (± 1 standard deviation (σ_{Δ}), with $\sigma_{\Delta} = \sqrt{(\sigma_{present\ day}^2 + \sigma_{2K}^2)/2}$). The red dots depict the ensemble mean response for 2019 and 2020.

A4.4b -- 2018 response to 2K warming

Next we examine the response of the hot and exceptionally dry growing season of 2018 that unfolded under persistent conditions of atmospheric blocking. In addition to the boxplots in Fig. 3, we show maps of the 2018 response anomaly with respect to the climatological mean response in Fig. 4 for 2018+2K|EC. Results for 2018+2K|HAD and MPI can be found in supplement C.

The response in winter and early spring preceding the blocking conditions is very similar to the climatological mean response in most variables and results in slightly wetter soil moisture conditions at the start of the growing season in April in 2018+2K|EC than in 2018REF. Also in spring, the 2018 soil moisture and circulation anomalies appear not to have a strong effect on the response. The precipitation response is – despite the application of the 11-member ensembles – rather patchy throughout the year (Fig. 4b). Consistent with the climatological mean response, precipitation is found to increase in April and May, with relatively strong increases in April. Apparently, the precipitation events originate from sources with sufficient moisture supply to sustain these increases. Evaporation increases with more than potential rate in these months, but only partially compensates the precipitation increase. Despite a small decrease in snowmelt and increase in runoff (not shown), the top 1 m of the soil is slightly wetter until mid-June in 2018+2K|EC than in 2018REF.

From mid-June onwards strong deviations from the climatological mean response occur for the temperature, relative humidity, atmospheric evaporative demand and evaporation responses, exceeding the 25th-75th percentile range (Fig. 3a,d,f,g and 4a,c,d).

Decreases in precipitation (June - September) and the weak evaporation response (July - September) show that sources of moisture are even more limited in a 2K warming scenario. Precipitation in this period is from predominantly continental sources (Benedict et al., 2021), and the evaporation response is moisture-constrained throughout Europe (Fig. 4c,d).

From July to September the temperature response over the basin area and surroundings is amplified with respect to the climatological mean response (+3.0K over JAS 2018 compared to +2.4K for the climatology). This response anomaly correlates with the anomalously low actual evaporation response. Within the basin, evaporation barely increases or even decreases in the period July - October, resulting in a further decrease in relative humidity, a modest increase in the sensible heat flux (Fig. 3c) and an increase in near-surface temperature. Note that the increase in solar radiation is relatively small in June and July (Fig. 3b), given the predominantly clear-sky conditions in 2018REF, and that the enhanced temperature response cannot be fully explained by the surface fluxes. Increases in heat advection due to stronger warming in upwind regions or enhanced warming through subsidence may play a role.

Since the response in summer evaporation in the west-central European river basins is close to zero, the JAS soil moisture response is small compared to most other years in the 1980-2020 period (Fig. 3h), and is almost completely determined by the precipitation decrease. The pattern of the soil moisture response anomaly strongly correlates with the precipitation response anomaly in this period (Fig. 4e). Percolation to deeper soil layers and runoff decrease in this period as well (not shown). In autumn and winter, moderate precipitation increases replenish the soils to 2018REF levels in December/January, in the top 1 m of the soil and in deeper layers.

A4.4c -- Sensitivity to the level of global warming

To examine the sensitivity of the results to the level of global warming, we additionally show the temperature and soil moisture conditions under +1.5K and +3K warming (Fig. 5). The temperature response is fairly linear with global warming under the increasingly moisture-constrained conditions, yielding temperature anomalies (i.e. deviations with respect to climREF) of +8.1K, +8.9K and +10.2K under respectively 1.5K, 2K and 3K warming during the July/August heat wave, and +4.8K, +5.4K and +6.5K for the growing season mean. The soil moisture depletion over the growing season decreases (increases) for lower (higher) levels of global warming, resulting in slightly less (more) intense drought conditions from mid-June onwards. The increase in soil moisture depletion over the growing season is mainly driven by stronger decreases in precipitation over JAS for higher levels of global warming, see Fig. 5g, in which we compare the hydrological budget changes over AMJ and JAS for the three warming levels. In AMJ, precipitation increases are weaker (stronger) under 1.5K (3K) warming, but so are the evaporation increase, snowmelt decrease and runoff increase, resulting in a slightly weaker soil moisture depletion over AMJ for higher levels of global warming. Although the JAS evaporative demand increases with higher levels of global warming, evaporation does not or only very weakly. In autumn and winter, stronger precipitation increases and less precipitation falling as snow under higher warming levels result in soil moisture replenishment to present-day levels in winter for all warming levels.

A4.4d -- Sensitivity to the perturbations

The results for the MPI-perturbed simulations are fairly similar in spring, but the soil moisture drying is stronger in summer and autumn, especially for 3K warming, owing to much stronger decreases in precipitation (Fig. 5c,d). JAS evaporation decreases, and the JAS temperature response is more strongly amplified compared to the climatological mean response than 2018+2K|EC.

The HAD-perturbed simulations (Fig. 5e,f) give a stronger soil moisture drying and near-surface heating early in the growing season, as we have seen for the climatological mean response (Sec. 4.1), but this is more pronounced under the circulation of 2018. This is associated with stronger decreases in precipitation and cloud cover in April and May and hence a stronger increase in solar radiation. The decline in the increase in evaporation that is common to all simulations occurs around a month earlier, and the sensible heat flux is enhanced compared to the climatological mean response throughout summer. The temperature response exceeds the climatological mean response throughout the growing season, and soils are drier than under present-day conditions from as early as the end of March until spring 2019.

A4.5 -- Impact on drought severity and frequency

A4.5a -- Increase in the 2018 drought severity

For all warming levels, the drought onset and ending under global warming occurs at roughly the same time as for the present-day event, with only slight differences between the individual members (Fig. 5b). The drought duration is thus hardly affected or even decreases compared to 2018REF, while the drought intensity shows a 22% increase (8 mm +/- 2 mm (mean +/- σ_{Δ})), resulting in an increase in the drought deficit volume of 20% under 2K warming. Tab. 2 summarizes the findings for all warming levels and GCMs. The HAD- and MPI-perturbed simulations show stronger increases in drought severity (+39%, resp. +25% under +2K warming) owing to decreasing precipitation in spring, respectively

weaker increases in precipitation in spring and stronger decreases in summer, especially under 3K warming. In general, drought intensity and severity increase with global warming, although drought episodes under 3K warming tend to be shorter than under 1.5K and 3K warming.

A4.5b -- Increase in severe drought occurrences under 2K warming

The boxplots in Fig. 3 reveal that there are summers in the 1980-2020 period with much stronger responses in the summer near-surface temperature and soil moisture than 2018. In Fig. 6 we show the co-occurrence of the July-September basin-mean soil moisture (SWI) and temperature response for each year in the 1980-2020 period, along with the absolute SWI and near-surface temperature under present-day and +2K conditions. The top 4 compound hot and dry years (2018, 2003, 2019 and 2020) in the present-day climate are marked, as well as the strongly responding years (1983 and 1989). The years with extreme soil moisture drying (1983, 1989) show strong warming, as expected, with a weaker correlation for smaller soil moisture responses. The soil moisture response for 2003, known for its severe hot and dry conditions and associated societal and economic impacts in central Europe (e.g. Rebetz et al., 2006, Fischer et al., 2007) is similar to the 2018 response, but with weaker warming. 2019 and 2020, similar under present-day conditions, respond very differently, with a small response for 2019, but relatively strong soil moisture and temperature responses for 2020. The ranking of compound hot and dry summers thus changes under +2K warming, 2018 still ranking first, but the deviation from other years decreases. 2003 is very similar to 1983, 2020 and 1989, with 1983 being the hottest, 1989 the driest summer after 2018.

The strongly responding summers (1983 and 1989) are moderately dry in the present-day climate (within the driest 25%), but years with soil moisture levels and temperatures that were very similar to these strongly responding years in the present-day climate, show much weaker responses. This suggests that large-scale circulation (variability) is still an important factor in the 'initialization' of the amplified soil moisture drying and warming. Under the specific circulation conditions in 1983, 1989 and 2020, the globally warmer background conditions result in strongly reduced precipitation, increased evaporation (except 1989) and soil drying in spring and early summer, resulting in a very strong response in incoming solar radiation, a negative response in evaporation and very strong increases in the sensible heat flux and near-surface temperatures later in summer.

To further quantify the impact of global warming on drought occurrences and severity in the west-central European river basins under +2K warming, we determine the basin-mean drought deficit volume, duration and intensity for all years in the 1980-2020 period under present-day and +2K conditions (Fig. 7). Again, the severity of the 2018 drought episode clearly stands out under present-day conditions. The drought deficit volume of 2003 is around three times smaller, with shorter duration, but especially smaller mean intensity. Furthermore, 2011, a spring drought (Trachte et al., 2012) that did not pup-up in the JAS soil moisture levels, and the 2019-2020 drought episodes stand out. Under +2K warming the drought frequency more than doubles compared to present-day conditions, reflecting the on average drier soil moisture conditions in summer and autumn. As we have seen for the JAS soil moisture response, the increase in drought severity is very strong for the years 1983 and 1989 (not qualified as drought under present-day levels), by far exceeding the present-day 2003 drought severity, and also the 2020 drought severity increases to above present-day 2003 levels. Note that these droughts have a stronger drought intensity, but that the duration is slightly shorter than the present-day 2003 drought event. The frequency of more severe droughts than the present-

day 2003 episode thus more than doubles, occurring on average once every eight years. This finding is confirmed by the results based on clim+2K MPI and clim+2K HAD, although amplitudes of the response slightly differ (Fig. C5 and C6). The April - March soil moisture evolution and spatial drought structure in JAS for the top 5 driest years under 2K warming is provided in Fig. 8. For 2018, apart from the intensification of the drought over the west- central European river basins, the future drought analogue affects a much larger part of Europe. The large spatial drought extent is a feature that is common to most future drought analogues. The drought intensification is not limited to central and southern Europe where the PGW simulations project the strongest soil moisture drying, but is seen for Sweden, Poland and the Baltic States as well.

A4.6 -- Discussion

In this study we have examined the implications of global warming for future drought severity in west-central Europe, by systematically perturbing the 1980-2020 period towards future climate conditions using the PGW approach. The control experiment has been carried out with the RCM KNMI-RACMO2 forced by large-scale information from the ERA5 reanalysis. The PGW-experiments use mean changes in temperature, humidity and winds derived from climate change information from GCM projections. Therewith results are less sensitive to GCM biases in the reference climate state, the atmospheric circulation is realistic per definition, the signal-to-noise ratio of the response is maximized, while a large part of the climate response is captured (Brogi et al., 2019b). Results can be directly related to real world events and their societal impact. This makes the results very tangible and therewith useful for climate change communication. Moreover, the PGW- simulations allow for a structural examination of the impact of adaptation measures or other changes in the system, and the PGW-simulations presented here have been used to investigate the impact of water withdrawals on freshwater availability (Greve et al., 2021), and ecosystem adaptation (Bouaziz et al., 2021) under climate change.

Here we have presented future analogues of the 2018 drought episode in west-central Europe, recorded as the most extreme drought in the 1980-2020 period in this region. It is generally hypothesized that under globally warmer conditions droughts set in earlier, last longer and are more intense, if conditions favoring a drought occur (Trenberth et al., 2014, Seneviratne et al., 2010). While this is true given equal precipitation deficits, here we show that for west-central Europe, under the anomalously persistent atmospheric blocking conditions of 2018, global warming yield increases in precipitation in spring, early summer and autumn that exceed the increase in evaporation, and prevent (EC and MPI-perturbed simulations) or limit (HAD-perturbed simulations) an increase in drought duration. The drought intensity moderately increases, resulting in an increase in soil moisture deficit volume of 20% (EC) to 39% (HAD) under +2K warming.

The intensification of the 2018 soil moisture drought under PGW is fairly small compared to the climatological mean soil moisture drying. The reason is that the drought evolution in 2018 is, already, strongly limited by the moisture-constrained conditions, and increases in precipitation in autumn terminate the 2018 drought episode. However, the temperature response is strongly amplified compared to the climatological mean response and relative humidity strongly decreases. Combined, this implies strong increases in stress on society, the natural environment and agriculture, and in wildfire risk (e.g. Krikken et al., 2019, Manning et al., 2019). More extreme drought conditions in terms of duration could of course occur under even more persistent or longer sequences of atmospheric

blocking conditions than during 2018, in years with drier antecedent winter and spring conditions, or, under stronger climate induced precipitation drying in spring than derived from the PGW experiments. Van der Wiel et al. (2021) sample drought events from a large ensemble of transient global climate model simulations (EC-EARTH) that match or exceed the 2018 drought conditions in the Rhine basin, and find indeed events with drier spring conditions than for the EC-perturbed simulations, but similar to the results based on the HAD-perturbed simulations.

Considering drought occurrences in the 1980-2020 period under 2K global warming, we have shown that the frequency of small to modest drought episodes strongly increases, as expected from the general soil moisture drying in (spring) summer and autumn, but more strikingly, that there is an unexpected increase in the frequency of 2003-like drought episodes owing to the occurrence of moderately dry years with very strong soil moisture drying and temperature responses. The non-linearity in the drying and warming response is related to the transition of energy-limited to moisture-limited evaporation regimes in the central-European regime transition area. Years with weakly moisture-limited evaporation under present-day conditions may respond strongly to decreases in precipitation and/or initially enhanced evaporation through shifts in the surface energy balance, further amplified by soil-moisture atmosphere feedbacks, in contrast to years such as 2018 and 2003 for which the response reaches a limit because these years are already in a moisture-constrained regime under present-day conditions (Schär et al., 2004, Lenderink et al., 2007, Seneviratne et al., 2010, Fischer et al., 2012). Moreover, the increased land-sea temperature gradient and south-north gradient over Europe may contribute to a stronger variability in heat and moisture advection (Lenderink et al., 2007, Holmes et al., 2016). Regarding the response in climatological mean soil moisture under PGW, our results indicate an increase in spring (HAD-perturbed simulations only), summer and autumn soil drying in the west-central European river basins under 2K global warming, yet soils are generally replenished to present-day levels in winter, in the top 1 m as well as in deeper soil layers. The general annual cycle in the soil moisture response is similar to results based on studies that use climate model ensembles of transient simulations, but e.g. Ruostenoja et al. (2018) and Van der Linden et al. (2019) find a soil moisture drying response in winter as well. Apart from potential differences owing to GCM structure and model resolution, the circulation component of climate change that is only partly taken into account in the PGW-approach thus likely contributes to a stronger soil moisture drying in spring and summer, for example through the projected higher frequency of easterly winds in spring and summer in west-central Europe bringing more often drier and warmer conditions (Haarsma et al., 2009, Van der Linden et al., 2019), or smaller increases in autumn and winter precipitation, as shown by Brogli et al. (2019b). With respect to change in the drought frequency, our results are within the range of drought frequency changes derived from transient climate model simulations, with a doubling of 2003-like soil moisture droughts under +3K warming found by Samaniego et al. (2018), and 2018-like drought conditions (SPEI) becoming the new normal within the second half of the 21st century found by Toreti et al. (2019). The increase in frequency in 2003-like droughts that we find based on the PGW-simulations is robust with respect to the selected GCM to derive the perturbations, and occurs independently of uncertain changes in the frequency of synoptic conditions that favor dry summers.

A4.7 -- Conclusion

Droughts and associated heat waves form a threat to society and nature, as again demonstrated in recent years in west-central Europe. To develop adaptation strategies, information of changes in drought risk under ongoing global warming is required. The genesis of soil moisture droughts is complex, but droughts generally develop under the influence of persistent anomalously high pressures, blocking the normal advection of moisture to west-central Europe. While there are indications that the frequency of blocking conditions increases in response to global warming, changes in dynamics are highly uncertain owing to natural variability and climate model biases. In this study we have examined the impact of global warming on the development and severity of a 2018-like drought, in relation to changes in drought severity in the 1980-2020 period, conditional on the historical circulation, i.e. we show what the 1980-2020 period would look like in a 2K warmer world. Therewith, changes in droughts can be directly related to events and their societal impact in the recent history, and the experiments provide tangible examples of what global warming entails.

Under 2K warming almost all years in the 1980-2020 period show a decrease in soil moisture availability in (spring), summer and autumn, consistent with results based on transient climate model simulations. The soil moisture response under the circulation of 2018 is fairly modest compared to the change in climatology, owing to the strong surface and canopy resistance to evaporation caused by low soil moisture availability and a strong vapor pressure deficit. The soil moisture deficit volume increases with 17% to 34% depending on warming level and GCM, owing to the increase in drought intensity. The drought duration is barely impacted owing to increasing precipitation in spring, autumn and winter. The increase in drought severity co-occurs with strong temperature increases.

We furthermore show that the response in temperature and soil moisture drying can be particularly large for years with moderately dry conditions in the present-day climate. Using present-day thresholds, the drought frequency strongly increases under +2K warming, with 2019/2020-like deficit volumes occurring on average every three years, 2003-like deficit volumes every eight years, and under strongly enhanced temperatures. This shows that even without taking into account changes in the frequency of atmospheric blocking conditions, the drought risk in west-central Europe is strongly enhanced.

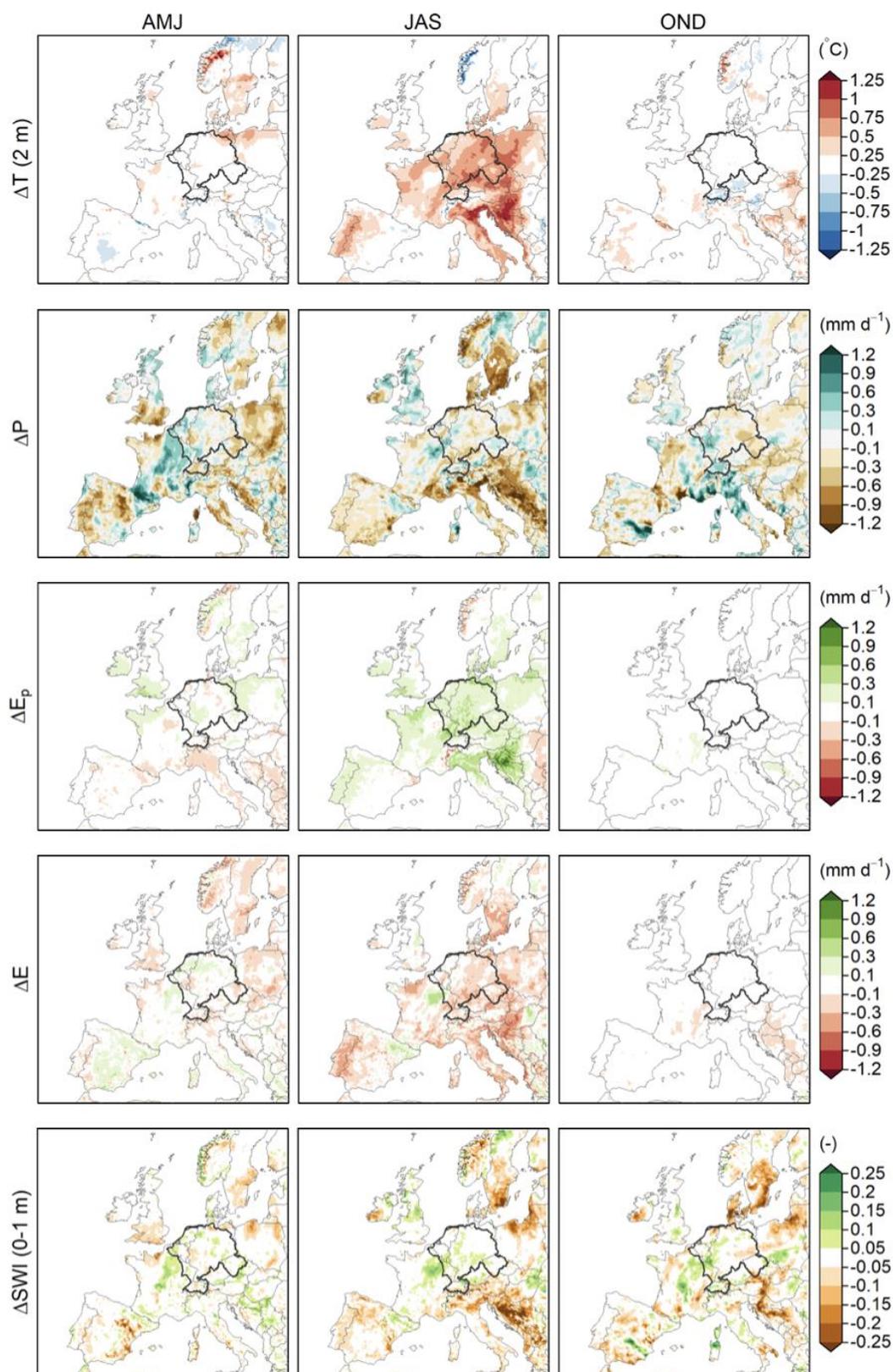


Figure 4: Difference between the 2018 and climatological mean (1980-2017) response to 2K warming (2018+2K EC - 2018REF) - (clim+2K EC-climREF) in a) temperature, b) precipitation, c,d) (potential) evaporation and e) SWI in the top 1 m of the soil. The response is masked over sea for clarity.

Table 2: Drought severity of the 2018 drought for present-day conditions (REF) and for +1.5K, +2K and +3K global warming. Listed are the ensemble mean (spread).

		Duration (d)		Intensity (mm)		Deficit volume (mm d)	
	REF	231	(11)	36	(3)	8238	(957)
	+1.5K	225	(12)	42	(2)	9408	(801)
EC	+2K	225	(11)	44	(1)	9899	(540)
	+3K	219	(11)	46	(2)	9998	(598)
	+1.5K	235	(1)	42	(1)	9928	(357)
MPI	+2K	236	(1)	44	(3)	10339	(622)
	+3K	226	(12)	48	(2)	10821	(639)
	+1.5K	252	(1)	45	(2)	11326	(605)
HA D	+2K	252	(1)	46	(1)	11480	(222)
	+3K	219	(19)	46	(1)	10144	(787)

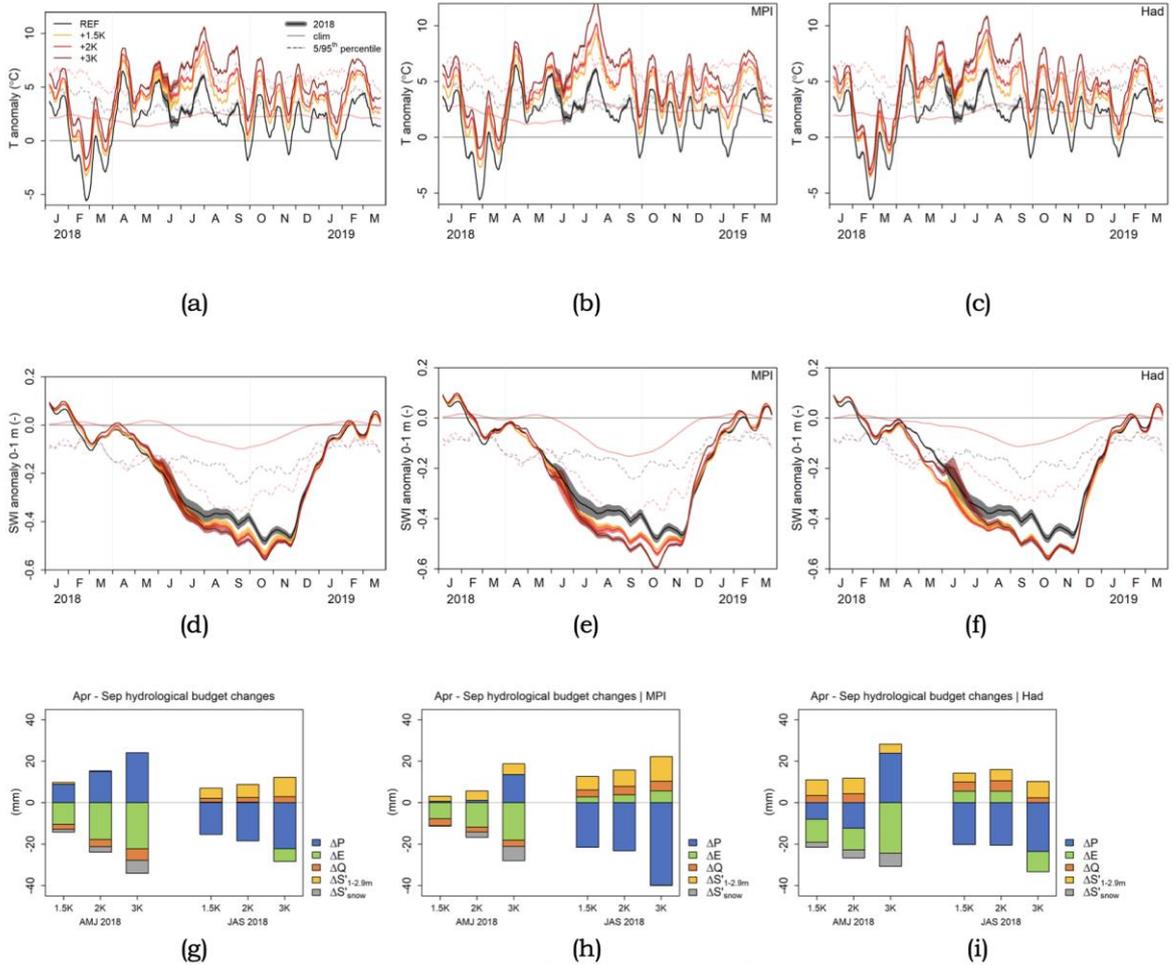


Figure 5: Basin-mean anomalies in (top row) the 2018 near-surface temperature and middle row) the SWI in the top 1m of the soil under present-day (grey), 1.5K (orange), 2K (red) and 3K (brown) global warming, and (bottom row) hydrological budget changes over AMJ and JAS 2018 in response to 1.5K, 2K and 3K global warming, for 2018+2K|EC (left column), 2018+2K|MPI (middle column) and 2018+2K|HAD (right column). All lines are determined as the deviation of the present-day climate mean (climREF). Solid lines with shading depict the 2018 ensemble mean and spread, as in Fig. 1; dashed lines the 95th (temperature) and 5th (SWI) percentile for present-day conditions and 2K warming; the solid line without shading depicts clim+2K – climREF, i.e. the climate response. ΔX = response in the seasonally integrated flux X (mm) or in the storage change X over the season (mm), with P = precipitation, E = evaporation, Q = runoff, $S'_{1-2.9m}$ = change in soil moisture storage in the bottom soil layer over the season and S'_{snow} = change in snow storage over the season, negative values indicating a decrease in snowmelt. The sum of the components equals the increase in soil moisture depletion in the top 1 m of the soil, meaning positive changes in precipitation are plotted on the positive y-axis, positive changes in all other variables are plotted on the negative y-axes.

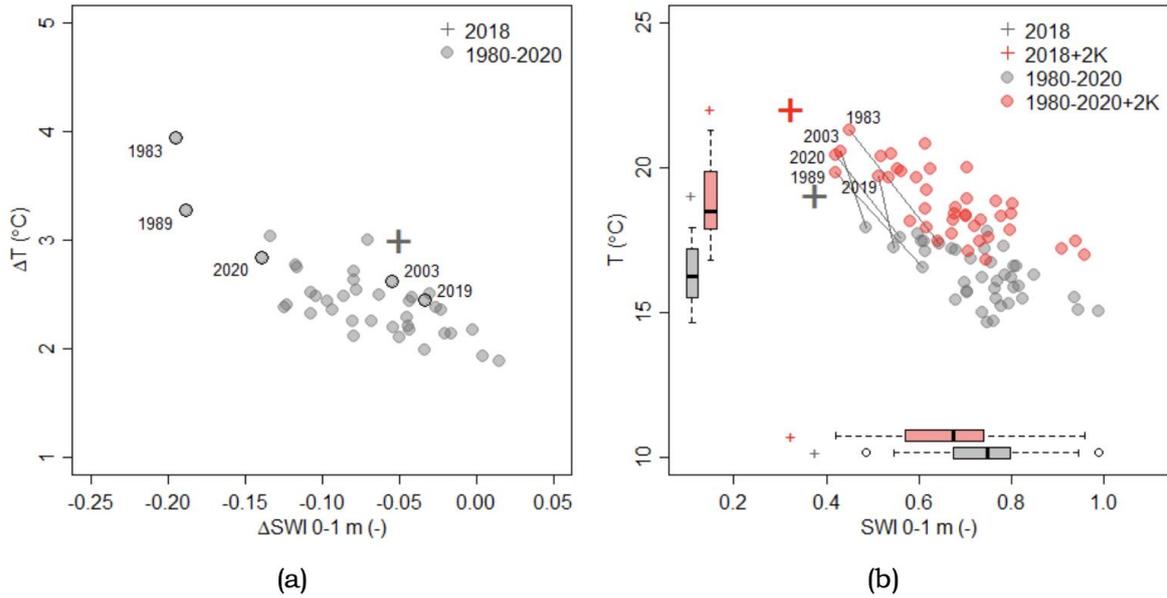


Figure 6: Scatterplot of the 1980-2020 JAS basin-mean near-surface temperature and SWI. (a) Response under 2K warming. (b) Absolute values under present-day (grey) and +2K (red) conditions and their distribution (boxplots). The +2K values are based on the EC-perturbed simulation. 2018 is indicated with a +. The strongly responding years and years with co-occurring low SWI values and high temperatures under present-day conditions are marked.

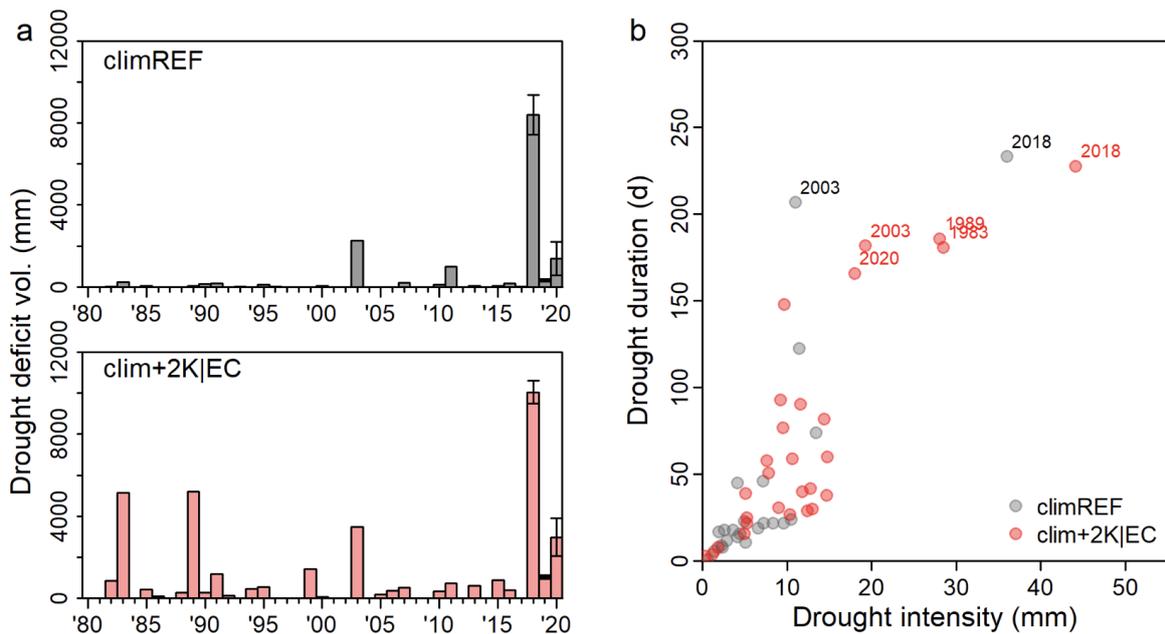


Figure 7: Drought severity in 1980-2020 under present-day (grey) and +2K conditions (red). Shown are a) the drought deficit volume, and b) drought duration and mean intensity. Droughts are determined for each hydrological year (April - March). The maximum drought per year is shown. Data for 2018-2020 is based on the 2018REF and +2K ensemble data. Error bars show the ensemble mean +/- 1 standard deviation.

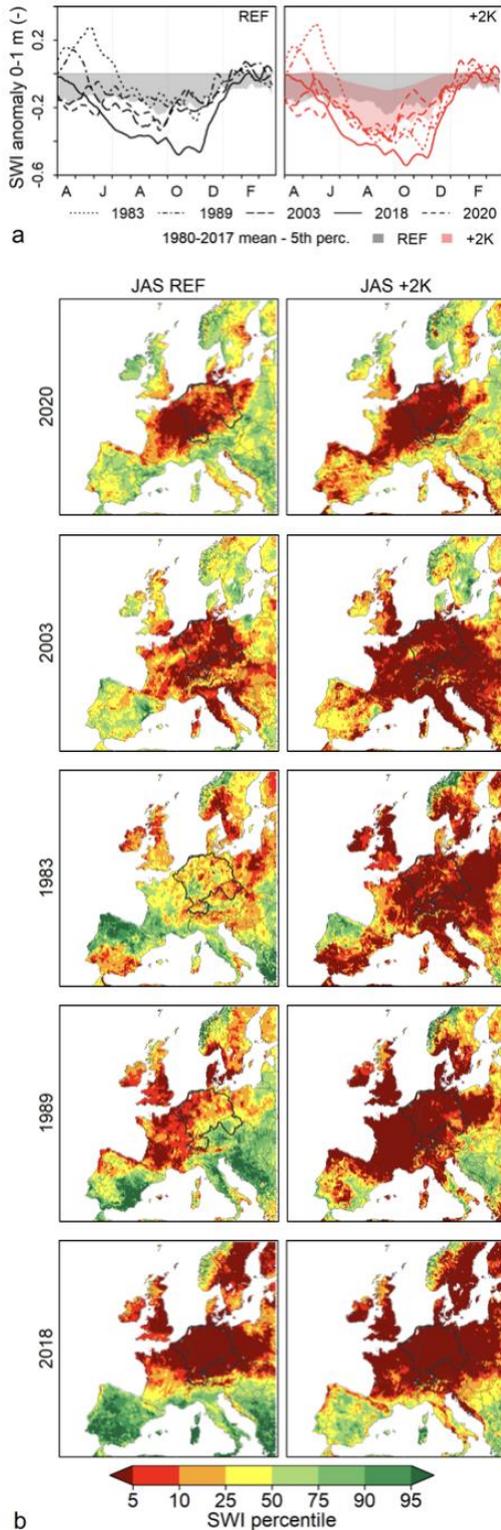


Figure 8: Top 5 droughts under +2K conditions, all exceeding the severity of the present-day 2003 drought in the west-central European river basins. a) Annual cycle of the SWI anomaly for present-day (left) and +2K conditions (right) as in Fig. 5, but with the 1980-2017 mean-5th percentile envelope shaded. b) Spatial structure of the top 5 droughts in JAS, showing the JAS-mean SWI percentile in the 1980-2017 period (climREF). Droughts are shown in order of increasing basin-mean drought severity under +2K warming. Dark red colors indicate severe drought conditions.

A4.8 -- References

- E. E. Aalbers, G. Lenderink, E. van Meijgaard, and B. J. van den Hurk. Local-scale changes in mean and heavy precipitation in western europe, climate change or internal variability? *Climate Dynamics*, 50(11):4745–4766, 2018.
- J. J. Attema, J. M. Loriaux, and G. Lenderink. Extreme precipitation response to climate perturbations in an atmospheric mesoscale model. *Environmental research letters*, 9 (1):014003, 2014.
- G. Balsamo, A. Beljaars, K. Scipal, P. Viterbo, B. van den Hurk, M. Hirschi, and A. K. Betts. A revised hydrology for the ecmwf model: Verification from field site to terrestrial water storage and impact in the integrated forecast system. *Journal of hydrometeorology*, 10(3):623–643, 2009. doi: <https://doi.org/10.1175/2008JHM1068.1>.
- D. Beillouin, B. Schauburger, A. Bastos, P. Ciais, and D. Makowski. Impact of extreme weather conditions on european crop production in 2018. *Philosophical Transactions of the Royal Society B*, 375(1810):20190510, 2020.
- I. Benedict, C. C. van Heerwaarden, E. C. van der Linden, A. H. Weerts, and W. Hazeleger. Anomalous moisture sources of the rhine basin during the extremely dry summers of 2003 and 2018. *Weather and Climate Extremes*, 31:100302, 2021.
- P. Bissolli. Europe and the middle east [in “state of the climate in 2018”]. *Bull. Am. Meteorol. Soc.*, 100:S217–S230, 2019. doi:10.1175/2019BAMSStateoftheClimate.1.
- L. J. Bouaziz, E. E. Aalbers, A. H. Weerts, M. Hegnauer, H. Buiteveld, R. Lammersen, J. Stam, E. Sprokkereef, H. H. Savenije, and M. Hrachowitz. The importance of ecosystem adaptation on hydrological model predictions in response to climate change. *Hydrology and Earth System Sciences Discussions*, pages 1–39, 2021.
- E. Brakkee, M. van Huijgevoort, and R. P. Bartholomeus. Spatiotemporal development of the 2018–2019 groundwater drought in the netherlands: a data-based approach. *Hydrology and Earth System Sciences Discussions*, pages 1–26, 2021.
- R. Brogli, S. L. Sørland, N. Kröner, and C. Schär. Causes of future mediterranean precipitation decline depend on the season. *Environmental Research Letters*, 14(11): 114017, 2019.
- R. Brogli, N. Kröner, S. L. Sørland, D. Lu`thi, and C. Schär. The role of hadley circulation and lapse-rate changes for the future european summer climate. *Journal of Climate*, 32(2):385–404, 2019b. doi: <https://doi.org/10.1175/JCLI-D-18-0431.1>.
- M. I. Brunner, K. Liechti, and M. Zappa. Extremeness of recent drought events in switzerland: dependence on variable and return period choice. *Natural Hazards and Earth System Sciences*, 19(10):2311–2323, 2019.
- M. Collins, R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichet, P. Friedlingstein, X. Gao, W. J. Gutowski, T. Johns, G. Krinner, et al. Long-term climate change: projections, commitments and irreversibility. In *Climate Change 2013-The Physical Science Basis: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 1029–1136. Cambridge University Press, 2013.
- W. Collins, N. Bellouin, M. Doutriaux-Boucher, N. Gedney, P. Halloran, T. Hinton, J. Hughes, C. Jones, M. Joshi, S. Liddicoat, et al. Development and evaluation of an earth-system model–hadgem2. *Geoscientific Model Development*, 4(4):1051–1075, 2011. doi: <https://doi.org/10.5194/gmd-4-1051-2011>.
- E. Coppola, R. Nogherotto, J. M. Ciarlo`, F. Giorgi, E. van Meijgaard, N. Kadygrov, C. Iles, L. Corre, M. Sandstad, S. Somot, et al. Assessment of the european climate projections as simulated by the large euro-cordex regional and global climate model ensemble. *Journal of Geophysical Research: Atmospheres*, 126(4):e2019JD032356, 2021.
- R. C. Cornes, G. van der Schrier, E. J. van den Besselaar, and P. D. Jones. An ensemble version of the e-obs temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres*, 123(17):9391–9409, 2018. doi:10.1029/2017JD028200
- D. Coumou, V. Petoukhov, S. Rahmstorf, S. Petri, and H. J. Schellnhuber. Quasi- resonant circulation regimes and hemispheric synchronization of extreme weather in boreal summer. *Proceedings of the National Academy of Sciences*, 111(34):12331– 12336, 2014.
- C. Deser, A. Phillips, V. Bourdette, and H. Teng. Uncertainty in climate change projections: the role of internal variability. *Climate Dynamics*, 38(3-4):527–546, 2010.
- ECMWF. *IFS Documentation CY33R1 - Part IV: Physical Processes*. Number 4 in IFS Documentation. ECMWF, 2009. doi: 10.21957/8o7vwlbdr. URL <https://www.ecmwf.int/node/9227>. ipz Operational implementation 3 June 2008i/pz.
- E. Fischer, J. Sedl`a`cek, E. Hawkins, and R. Knutti. Models agree on forced response pattern of precipitation and temperature extremes. *Geophysical Research Letters*, 41 (23):8554–8562, 2014.
- E. M. Fischer, S. I. Seneviratne, P. L. Vidale, D. Lu`thi, and C. Schär. Soil moisture– atmosphere interactions during the 2003 european summer heat wave. *Journal of Climate*, 20(20):5081–5099, 2007.
- E. M. Fischer, J. Rajczak, and C. Schär. Changes in european summer temperature variability revisited. *Geophysical Research Letters*, 39(19), 2012.
- M. A. Giorgetta, J. Jungclaus, C. H. Reick, S. Legutke, J. Bader, M. B`ottinger, V. Brovkin, T. Crueger, M. Esch, K. Fieg, et al. Climate and carbon cycle changes from 1850 to 2100 in mpi-esm simulations for the coupled model intercomparison project phase 5. *Journal of Advances in Modeling Earth Systems*, 5(3):572–597, 2013. doi:10.1002/jame.20038.
- P. Greve, P. Burek, L. Guillaumot, E. van Meijgaard, E. Aalbers, M. Smilovic, S.-W. F., T. Kahil, and Y. Wada. Low flow sensitivity to water withdrawals under 2k global warming. *subm.*, 2021.

- L. Gudmundsson and S. I. Seneviratne. Anthropogenic climate change affects meteorological drought risk in Europe. *Environmental Research Letters*, 11(4):044005, 2016.
- R. J. Haarsma, F. Selten, B. V. Hurk, W. Hazeleger, and X. Wang. Drier mediterranean soils due to greenhouse warming bring easterly winds over summertime central Europe. *Geophysical research letters*, 36(4), 2009.
- M. Hanel, O. Rakovec, Y. Markonis, P. M'aca, L. Samaniego, J. Kysely, and R. Kumar. Revisiting the recent European droughts from a long-term perspective. *Scientific reports*, 8(1):1–11, 2018. doi: <https://doi.org/10.1038/s41598-018-27464-4>.
- V. Hari, O. Rakovec, Y. Markonis, M. Hanel, and R. Kumar. Increased future occurrences of the exceptional 2018–2019 central European drought under global warming. *Scientific reports*, 10(1):1–10, 2020.
- W. Hazeleger, C. Severijns, T. Semmler, S. Stefanescu, S. Yang, X. Wang, K. Wyser, E. Dutra, J. M. Baldasano, R. Bintanja, et al. Earth. *Bulletin of the American Meteorological Society*, 91(10):1357, 2010.
- W. Hazeleger, B. J. van den Hurk, E. Min, G. J. van Oldenborgh, A. C. Petersen, D. A. Stainforth, E. Vasileiadou, and L. A. Smith. Tales of future weather. *Nature Climate Change*, 5(2):107–113, 2015. doi: <https://doi.org/10.1038/nclimate2450>.
- H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Hor'anyi, J. Mun'oz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, et al. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020.
- C. R. Holmes, T. Woollings, E. Hawkins, and H. De Vries. Robust future changes in temperature variability under greenhouse gas forcing and the relationship with thermal advection. *Journal of Climate*, 29(6):2221–2236, 2016.
- D. Jacob, J. Petersen, B. Eggert, A. Alias, O. B. Christensen, L. M. Bouwer, A. Braun, A. Colette, M. D'equ'e, G. Georgievski, et al. Euro-cordex: new high-resolution climate change projections for European impact research. *Regional Environmental Change*, 14(2):563–578, 2014.
- P. Jarvis. The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 273(927):593–610, 1976.
- K. Kornhuber, S. Osprey, D. Coumou, S. Petri, V. Petoukhov, S. Rahmstorf, and L. Gray. Extreme weather events in early summer 2018 connected by a recurrent hemispheric wave-7 pattern. *Environmental Research Letters*, 14(5):054002, 2019. doi: <https://doi.org/10.1088/1748-9326/ab13bf>.
- N. Kramer, M. Mens, J. Beersma, and N. Kielen. Hoe extreem was de droogte van 2018?, 2019.
- F. Krikken, F. Lehner, K. Haustein, I. Drobyshev, and G. J. van Oldenborgh. Attribution of the role of climate change in the forest fires in Sweden 2018. *Natural Hazards and Earth System Sciences Discussions*, pages 1–24, 2019.
- N. Kr'oner, S. Kotlarski, E. Fischer, D. Lu'thi, E. Zubler, and C. Sch'ar. Separating climate change signals into thermodynamic, lapse-rate and circulation effects: theory and application to the European summer climate. *Climate Dynamics*, 48(9-10):3425–3440, 2017.
- A. La'in'e, H. Nakamura, K. Nishii, and T. Miyasaka. A diagnostic study of future evaporation changes projected in CMIP5 climate models. *Climate dynamics*, 42(9-10):2745–2761, 2014.
- G. Lenderink, A. Van Ulden, B. Van den Hurk, and E. Van Meijgaard. Summertime inter-annual temperature variability in an ensemble of regional model simulations: analysis of the surface energy budget. *Climatic Change*, 81(1):233–247, 2007.
- G. Lenderink, D. Belu'si'c, H. J. Fowler, E. Kjellstr'om, P. Lind, E. van Meijgaard, B. van Ulfst, and H. de Vries. Systematic increases in the thermodynamic response of hourly precipitation extremes in an idealized warming experiment with a convection-permitting climate model. *Environmental Research Letters*, 14(7):074012, 2019.
- C. Manning, M. Widmann, E. Bevacqua, A. F. Van Loon, D. Maraun, and M. Vrac. Increased probability of compound long-duration dry and hot events in Europe during summer (1950–2013). *Environmental Research Letters*, 14(9):094006, 2019. doi: <https://doi.org/10.1088/1748-9326/ab23bf>.
- R. Orth, J. Zscheischler, and S. I. Seneviratne. Record dry summer in 2015 challenges precipitation projections in central Europe. *Scientific Reports*, 6(1):1–8, 2016. doi: <https://doi.org/10.1038/srep28334>.
- R. S. Padr'on, L. Gudmundsson, B. Decharme, A. Ducharme, D. M. Lawrence, J. Mao, D. Peano, G. Krinner, H. Kim, and S. I. Seneviratne. Observed changes in dry-season water availability attributed to human-induced climate change. *Nature geoscience*, 13(7):477–481, 2020.
- S. Y. Philip, S. F. Kew, K. van der Wiel, N. Wanders, and G. J. van Oldenborgh. Regional differentiation in climate change induced drought trends in the Netherlands. *Environmental Research Letters*, 15(9):094081, 2020.
- A. F. Prein, R. M. Rasmussen, K. Ikeda, C. Liu, M. P. Clark, and G. J. Holland. The future intensification of hourly precipitation extremes. *Nature Climate Change*, 7(1): 48–52, 2017. doi: <https://doi.org/10.1038/nclimate3168>.
- M. Rebetez, H. Mayer, O. Dupont, D. Schindler, K. Gartner, J. P. Kropp, and A. Menzel. Heat and drought 2003 in Europe: a climate synthesis. *Annals of Forest Science*, 63(6):569–577, 2006.
- B. R'osner, I. Benedict, C. van Heerwaarden, and et al. Sidebar 7.3: The long heat wave and drought in Europe in 2018 [in "State of the climate in 2018"]. *Bull. Am. Meteorol. Soc.*, 100(9):S221–S237, 2019. doi: [doi:10.1175/2019BAMSStateoftheClimate.1](https://doi.org/10.1175/2019BAMSStateoftheClimate.1).
- K. Ruosteenoja, T. Markkanen, A. Ven'al'ainen, P. R'ais'anen, and H. Peltola. Seasonal soil moisture and drought occurrence in Europe in CMIP5 projections for the 21st century. *Climate Dynamics*, 50(3):1177–1192, 2018. doi: <https://doi.org/10.1007/s00382-017-3671-4>.

- L. Samaniego, S. Thober, R. Kumar, N. Wanders, O. Rakovec, M. Pan, M. Zink, J. Sheffield, E. F. Wood, and A. Marx. Anthropogenic warming exacerbates european soil moisture droughts. *Nature Climate Change*, 8(5):421–426, 2018.
- C. Schär, C. Frei, D. Lüthi, and H. C. Davies. Surrogate climate-change scenarios for regional climate models. *Geophysical Research Letters*, 23(6):669–672, 1996.
- C. Schär, P. L. Vidale, D. Lüthi, C. Frei, C. H. Aberli, M. A. Liniger, and C. Appenzeller. The role of increasing temperature variability in european summer heatwaves. *Nature*, 427(6972):332–336, 2004.
- B. Schuldt, A. Buras, M. Arend, Y. Vitasse, C. Beierkuhnlein, A. Damm, M. Gharun, T. E. Grams, M. Hauck, P. Hajek, et al. A first assessment of the impact of the extreme 2018 summer drought on central european forests. *Basic and Applied Ecology*, 45:86–103, 2020.
- F. M. Selten, R. Bintanja, R. Vautard, and B. van den Hurk. Future continental summer warming constrained by the present-day seasonal cycle of surface hydrology. *Scientific reports*, 10(1):1–7, 2020.
- S. Seneviratne, N. Nicholls, D. Easterling, C. Goodess, S. Kanae, J. Kossin, Y. Luo, J. Marengo, K. McInnes, M. Rahimi, et al. Managing the risks of extreme events and disasters to advance climate change adaptation. a special report of working groups i and ii of the intergovernmental panel on climate change (ipcc), 2012.
- S. I. Seneviratne, T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling. Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, 99(3-4):125–161, 2010.
- T. G. Shepherd. Atmospheric circulation as a source of uncertainty in climate change projections. *Nature Geoscience*, 7(10):703–708, 2014.
- T. G. Shepherd. Storyline approach to the construction of regional climate change information. *Proceedings of the Royal Society A*, 475(2225):20190013, 2019.
- J. Sillmann, T. Shepherd, B. van den Hurk, W. Hazeleger, O. Romppainen-Martius, and J. Zscheischler. Physical modeling supporting a storyline approach. *CICERO Policy Note*, 2019(1), 2019.
- R. Sluijter, M. Plieger, G. Van Oldenborgh, J. Beersma, and H. De Vries. De droogte van een analyse op basis van het potentiële neerslagtekort. *Tech. rep. Royal Netherlands Meteorological Institute*, 2018.
- P. M. Sousa, R. M. Trigo, D. Barriopedro, P. M. Soares, A. M. Ramos, and M. L. Liberato. Responses of european precipitation distributions and regimes to different blocking locations. *Climate dynamics*, 48(3):1141–1160, 2017.
- P. M. Sousa, R. M. Trigo, D. Barriopedro, P. M. Soares, and J. A. Santos. European temperature responses to blocking and ridge regional patterns. *Climate Dynamics*, 50(1):457–477, 2018.
- P. A. Stott, D. A. Stone, and M. R. Allen. Human contribution to the european heatwave of 2003. *Nature*, 432(7017):610–614, 2004.
- K. E. Taylor, R. J. Stouffer, and G. A. Meehl. An overview of cmip5 and the experiment design. *Bulletin of the American meteorological Society*, 93(4):485–498, 2012.
- A. Toreti, A. Belward, I. Perez-Dominguez, G. Naumann, J. Luterbacher, O. Cronie, L. Seguini, G. Manfron, R. Lopez-Lozano, B. Baruth, et al. The exceptional 2018 european water seesaw calls for action on adaptation. *Earth's Future*, 7(6):652–663, 2019.
- K. Trachte, A. Obregón, P. Bissolli, J. J. Kennedy, D. E. Parker, R. M. Trigo, and D. Barriopedro. Europe [in “state of the climate in 2011”]. *Bull. Am. Meteorol. Soc.*, 93:S186–S191, 2012. doi: doi:10.1175/2019BAMSStateoftheClimate.1.
- K. E. Trenberth, A. Dai, G. Van Der Schrier, P. D. Jones, J. Barichivich, K. R. Briffa, and J. Sheffield. Global warming and changes in drought. *Nature Climate Change*, 4(1):17–22, 2014.
- E. C. Van der Linden, R. J. Haarsma, and G. v. d. Schrier. Impact of climate model resolution on soil moisture projections in central-western europe. *Hydrology and Earth System Sciences*, 23(1):191–206, 2019.
- K. van der Wiel, G. Lenderink, and H. de Vries. Physical storylines of future european drought events like 2018 based on ensemble climate modelling. *Weather and Climate Extremes*, page 100350, 2021.
- E. Van Meijgaard, L. van Ulft, G. Lenderink, S. de Roode, E. Wipfler, R. Boers, and R. van Timmermans. Refinement and application of a regional atmospheric model for climate scenario calculations of western europe, 2012.
- R. Vautard, M. van Aalst, O. Boucher, A. Drouin, K. Haustein, F. Kreienkamp, G. J. Van Oldenborgh, F. E. Otto, A. Ribes, Y. Robin, et al. Human contribution to the record-breaking june and july 2019 heatwaves in western europe. *Environmental Research Letters*, 15(9):094077, 2020.
- M. M. Vogel, J. Zscheischler, and S. I. Seneviratne. Varying soil moisture–atmosphere feedbacks explain divergent temperature extremes and precipitation projections in central europe. *Earth System Dynamics*, 9(3):1107–1125, 2018.
- M. M. Vogel, J. Zscheischler, R. Wartenburger, D. Dee, and S. I. Seneviratne. Concurrent 2018 hot extremes across northern hemisphere due to human-induced climate change. *Earth's future*, 7(7):692–703, 2019.
- K. Wehrli, M. Hauser, and S. I. Seneviratne. Storylines of the 2018 northern hemisphere heatwave at pre-industrial and higher global warming levels. *Earth System Dynamics*, 11(4):855–873, 2020.
- T. Woollings, D. Barriopedro, J. Methven, S.-W. Son, O. Martius, B. Harvey, J. Sillmann, A. R. Lupo, and S. Seneviratne. Blocking and its response to climate change. *Current Climate Change Reports*, 4(3):287–300, 2018.
- P. Yiou, J. Cattiaux, D. Faranda, N. Kadyrov, A. Jézéquel, P. Naveau, A. Ribes, Y. Robin, S. Thao, G. J. van Oldenborgh, et al. Analyses of the northern european summer heatwave of 2018. *Bulletin of the American Meteorological Society*, 101(1): S35–S40, 2020.

J. Zscheischler and E. M. Fischer. The record-breaking compound hot and dry 2018 growing season in germany. *Weather and Climate Extremes*, 29:100270, 2020. ISSN 2212-0947. doi:<https://doi.org/10.1016/j.wace.2020.100270>.

Appendix

A5 - Appendix to Science part 7 - Selecting weather events from initial condition large ensembles (KNMI)

The science highlight described in the main deliverable is a summary of a full study. More details can be found in the peer reviewed, open access study:

K van der Wiel, G Lenderink, H de Vries (2021): Physical storylines of future European drought events like 2018 based on ensemble climate modelling. *Weather and Climate Extremes*, 33, pp. 100350.

Online available at: <http://doi.org/10.1016/j.wace.2021.100350>